

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuLink Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:22:15, Search time 58.84 seconds
(without alignments)
3.735 Million cell updates/s

Title: US-09-744-804-35

Perfect score: 53

Sequence: 1 KQCFENAW 9

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 231528 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231528

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

Issued_Patents_AA:*

- 1: /seq2/6/prodata/2/seq2/AA/MR/PF *
- 2: /seq2/6/prodata/2/seq2/AA/MR/PF *
- 3: /seq2/6/prodata/2/seq2/AA/MR/PF *
- 4: /seq2/6/prodata/2/seq2/AA/MR/PF *
- 5: /seq2/6/prodata/2/seq2/AA/MR/PF *
- 6: /seq2/6/prodata/2/seq2/AA/MR/PF *

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	53	100.0	159	2	US-08-162-402B-12
2	53	100.0	177	1	US-07-607-536C-3
3	53	100.0	217	2	US-08-162-402B-3
4	53	100.0	218	1	US-07-607-538C-2
5	53	100.0	218	2	US-08-162-402B-2
6	53	100.0	387	2	US-08-162-402B-6
7	53	100.0	445	2	US-08-162-402B-8
8	39	73.5	359	2	US-08-424-224-2
9	39	73.5	369	5	PCT-US94/02891-69
10	36	67.5	421	2	US-08-480-229C-21
11	36	67.5	421	2	US-08-480-229C-21
12	36	67.9	480	2	US-08-480-229C-10
13	36	67.9	480	2	US-08-480-229C-10
14	36	67.9	513	2	US-08-480-229C-14
15	34	67.9	513	2	US-08-480-229C-14
16	34	64.2	347	1	US-08-052-205-7
17	34	64.2	347	1	US-08-595-974-7
18	34	64.2	369	1	US-08-052-205-4
19	34	64.2	369	1	US-08-595-974-4
20	33	62.3	227	1	US-08-452-023B-4
21	33	62.3	229	4	US-08-121-436A-4
22	33	62.3	1841	3	US-08-908-978-2
23	32	60.4	157	2	US-08-162-402B-13
24	32	60.4	340	2	US-08-480-229C-20
25	32	60.4	329	2	US-08-659-236C-20
26	32	60.4	444	2	US-08-162-402B-62
27	32	60.4	444	2	US-08-162-402B-6

28	32	60.4	477	1	US-08-191-338A-2
29	32	60.4	1121	1	US-07-799-915A-2
30	32	60.4	1121	1	US-08-335-522C-2
31	32	60.4	1121	1	US-08-487-203A-2
32	32	60.4	1334	6	5476657-1
33	32	60.4	1403	1	US-07-908-253-3
34	32	60.4	1403	2	US-08-694-865-17
35	32	60.4	1403	2	US-08-694-865-17
36	32	60.4	1403	3	US-09-124-491-17
37	32	60.4	1541	4	US-08-296-741-3
38	32	60.4	1541	5	PCT-US95-10661A-3
39	32	60.4	1545	4	US-08-296-791-4
40	32	60.4	1545	5	PCT-US95-10661A-4
41	32	60.4	1702	4	US-08-296-791-5
42	32	60.4	1702	5	PCT-US95-10661A-5
43	31	58.5	107	1	US-08-036-210-4
44	31	58.5	107	2	US-08-449-609-4
45	31	58.5	159	2	US-08-162-402B-16

ALIGNMENTS

RESULT 1

US-08-162-402B-12

: Sequence 12, Application US/08152402B

: Patent No. 5972437

: GENERAL INFORMATION:

: APPLICANT: CERTANI, ROBERTO L.

: APPLICANT: PETERSON, JERRY A.

: APPLICANT: LAPATCA, DAVID J.

: TITLE: A METHOD FOR THE PREPARATION OF A RECOMBINANT HUMAN MILK FAT

: NUMBER OF SEQUENCES: 29

: CORRESPONDING ADDRESS:

: ADDRESS: Pretty, Schroeder & Poplawski

: STREET: 444 South Flower St., 19th Floor

: CITY: Los Angeles

: STATE: CA

: COUNTRY: USA

: ZIP: 90071

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Diskette

: COMPUTER: IBM Compatible

: OPERATING SYSTEM: DOS

: SOFTWARE: FASTSEQ for Windows Version 2.0

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/162402B

: FILING DATE: 03/02/1993

: CLASSIFICATION: 435

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER:

: FILING DATE:

: ATTORNEY/AGENT INFORMATION:

: NAME: Amel, Virginia

: REGISTRATION NUMBER: 30,940

: REFERENCE/SCRIPT NUMBER: PCT 88215

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 213-489-7799

: TELEFAX: 213-489-4210

: TITEL

: INFORMATION FOR SEQ ID NO: 12:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 159 amino acids

: TYPE: amino acid

: STEADINESS: unknown

: TOPOLOGY: unknown

: MOTIF TYPE: peptide

: US 08-162-402B-12

Query Match 100.0%, Score 53, DB 2, Length 159,

Best Local Similarity 100.0%, Prod. No. 0.034:

```

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KCGNFNAW 9
   |||||
Db 43 KCGNFNAW 51

RESULT 2
US-07-607-538C-4
; Sequence 3, Application US/07607538C
; Patent No. 5455031
; GENERAL INFORMATION:
; APPLICANT: Ceriani Dr., Roberto L.
; APPLICANT: Peterson Dr., Jerry A.
; TITLE OF INVENTION: POLYPEPTIDE WITH 46
; TITLE OF INVENTION: DIFFERENTIATION ANTIGEN BINDING SPECIFICITY AND CLOTTING
; TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
; TITLE OF INVENTION: FUSION PROTEIN, POLYPEPTIDE AND POLYPEPTIDE-
; TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-
; TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: V. Anzel & Assoc.
; STREET: 2055 No. 5455031th Broadway
; CITY: Walnut Creek
; STATE: California
; COUNTRY: USA
; ZIP: 94596
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 5.0
; SOFTWARE: Patent In Release #1.6, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/607,538C
; FILING DATE: 01 NOV 1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Viviana Anzel
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: CRPGC-004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 943-1931
; TELEFAX: (510) 943-1189
; TELEX: N.A.
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 217 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE:
US-07-607-538C-4

Query Match 100.0%, Score 53, DB 1, Length 217,
Best Local Similarity 100.0%, Pred. No. 0.047,
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KCGNFNAW 9
   |||||
Db 101 KCGNFNAW 109

RESULT 4
US-08-162-402B-3
; Sequence 4, Application US/08162402B
; Patent No. 5972347
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.

```

```

; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: 46 KDA HUMAN MILK FAT
; TITLE OF INVENTION: GLOBULIN (HMPG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Anzel, Viviana
; REGISTRATION NUMBER: 30,940
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 217 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-162-402B-3

Query Match 100.0%, Score 53, DB 2, Length 217,
Best Local Similarity 100.0%, Pred. No. 0.047,
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KCGNFNAW 9
   |||||
Db 101 KCGNFNAW 109

RESULT 4
US-07-607-538C-2
; Sequence 2, Application US/07607538C
; Patent No. 5455031
; GENERAL INFORMATION:
; APPLICANT: Ceriani Dr., Roberto L.
; APPLICANT: Peterson Dr., Jerry A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: POLYPEPTIDE WITH 46
; TITLE OF INVENTION: DIFFERENTIATION ANTIGEN BINDING SPECIFICITY AND CLOTTING
; TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
; TITLE OF INVENTION: FUSION PROTEIN, POLYPEPTIDE AND POLYPEPTIDE-
; TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-
; TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: V. Anzel & Assoc.
; STREET: 2055 No. 5455031th Broadway
; CITY: Walnut Creek
; STATE: California
; COUNTRY: USA

```

ZIP: 94596
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 5.0
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/607,538C
FILING DATE: 01-NOV-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Viviana Amzel
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: CPCC-004
TELEPHONE: (510) 943-1941
TELEFAX: (510) 943-1189
TELEX: N A
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE:
US-07-607-538C-2

Query Match 100.0%; Score 53; DB 1; Length 218;
Best Local Similarity 100.0%; Pred No. 0.847;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KOGNFNAWV 9
|||||
DB 102 KOGNFNAWV 110

RESULT 5
US-08-162-402B-2
Sequence 2, Application US/08162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERIANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LARocca, DAVID J.
TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
TITLE OF INVENTION: GLOBULIN (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplawski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162,402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: PC 38215
TELECOMMUNICATION INFORMATION:

TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-09-162-402B-2

Query Match 100.0%; Score 53; DB 2; Length 218;
Best Local Similarity 100.0%; Pred No. 0.047;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KOGNFNAWV 9
|||||
DB 102 KOGNFNAWV 110

RESULT 6
US-08-162-402B-6
Sequence 6, Application US/08162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERIANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LARocca, DAVID J.
TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
TITLE OF INVENTION: GLOBULIN (HMFG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplawski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162,402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: PC 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 387 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-6

Query Match 100.0%; Score 53; DB 2; Length 387;
Best Local Similarity 100.0%; Pred No. 0.084;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQGNFNAW 9
|||||
DB 271 KQGNFNAW 279

RESULT 7

US-08-162-402B-#
Sequence 8, Application US/08162402B
Patent No. 5972347
GENERAL INFORMATION:
APPLICANT: CERIANT, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LAROCCA, DAVID J.
TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
TITLE OF INVENTION: GLOBULE (HMPG) ANTIGEN, FRAGMENTS & PEPTIDE PROTEIN
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Betty Schaefer & Papiowski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US-08/162-402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: P66 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7780
TELEFAX: 213-489-4210
TELEX:
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 465 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-#

Query Match 100.0%; Score 53; DB 2; Length 465,
Best Local Similarity 100.0%; Pred. No. 0.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KQGNFNAW 9
|||||
DB 349 KQGNFNAW 357

RESULT 8

US-08-424-224-2
Sequence 2, Application US/08424224
Patent No. 5912173
GENERAL INFORMATION:
APPLICANT: LEONARD, WARREN J.
TITLE OF INVENTION: MARINE IL-2R CDNA AND
TITLE OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVE.
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORD PERFECT # 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/424,224
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/121,435
FILING DATE: 14-SEPT-1993
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAM S. FEILER
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-40610S1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: 212-751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 369
TYPE: AMINO ACID
TOPOLOGY: UNKNOWN
MOLECULE TYPE: PROTEIN
DESCRIPTION: NO
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: MURINE
INDIVIDUAL ISOLATE: IL-2R
US 08 424 224-2

Query Match 73.5%; Score 29; DB 2; Length 369;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QGNFNAW 8
|||||
DB 304 QGNFSAW 310

RESULT 9

PCI-US94-02891-69
Sequence 69, Application PC/TUS9402891
GENERAL INFORMATION:
APPLICANT: THE GOVERNMENT OF THE UNITED STATES OF AMERICA AS
APPLICANT: REPRESENTED BY THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN
APPLICANT: SERVICES
APPLICANT: OFFICE OF TECHNOLOGY TRANSFER, NATIONAL
APPLICANT: INSTITUTES OF HEALTH, BOX OTT, BETHESDA, MARYLAND 20892 USA
TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT OF
TITLE OF INVENTION: XSCID
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVE.
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORD PERFECT # 5.1

1 CURRENT APPLICATION DATA:
2 APPLICATION NUMBER: 08/031,143
3 FILING DATE: 12-MAR-1993
4 CLASSIFICATION: 11
5 PRIOR APPLICATION NUMBER: 08/031,143
6 FILING DATE: 12-MAR-1993
7 APPLICATION NUMBER: 08/121,435
8 FILING DATE: 14-SEP-1993
9 ATTORNEY/AGENT INFORMATION:
10 NAME: WILLIAM S. FRIED
11 REGISTRATION NUMBER: 26,728
12 REFERENCE/POCKET NUMBER: 8907-0026-999
13 TELECOMMUNICATION INFORMATION:
14 TELEPHONE: 212-758-4800
15 TELEFAX: 212-751-6849
16 TELETYPE: 421792
17 INFORMATION FOR SEQ ID NO: 69:
18 SEQUENCE CHARACTERISTICS:
19 LENGTH: 369
20 TYPE: AMINO ACID
21 TOPOLOGY: UNKNOWN
22 MOLECULE TYPE: PROTEIN
23 DESCRIPTION: NO
24 HYPOTHETICAL: NO
25 ORIGINAL SOURCE:
26 ORGANISM: MURINE
27 INDIVIDUAL ISOLATE: IL-2R
28 PCT-US94-02891-69

Query Match: 73.6%; Score 39; DB 5; Length 369;
Best Local Similarity: 85.7%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QGNFNW 8

DB 304 QGNFSW 310

RESULT 10
US-08-480-229C-21
1 Sequence 21, Application: US/08480229C
2 Patent No. 5874562
3 GENERAL INFORMATION:
4 APPLICANT: Quertemous, Thomas
5 APPLICANT: Hogan, Bridgid
6 APPLICANT: Snodgrass, H. Ralph
7 APPLICANT: Zupancic, Thomas J.
8 TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
9 NUMBER OF SEQUENCES: 29
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: Pennie & Edmonds LLP
12 CITY: New York
13 STATE: New York
14 COUNTRY: United States
15 ZIP: 10036 2711
16 COMPUTER READABLE FORM:
17 OPERATING SYSTEM: PC-DOS/MS-DOS
18 SOFTWARE: Patent in Release #1.0, Version #1.30
19 CURRENT APPLICATION DATA:
20 FILING DATE: 07-JUN-1995
21 CLASSIFICATION: 536
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Poissant, Brian M.
24 REGISTRATION NUMBER: 29,462
25 REFERENCE/POCKET NUMBER: 8907-0026-999
26 TELECOMMUNICATION INFORMATION:

1 TELEPHONE: (212) 790-9090
2 TELEFAX: (212) 869-8854/9741
3 INFORMATION FOR SEQ ID NO: 21:
4 SEQUENCE CHARACTERISTICS:
5 LENGTH: 321 amino acids
6 TYPE: amino acid
7 STRANDEDNESS:
8 TOPOLOGY: unknown
9 MOLECULE TYPE: protein
10 US-08-480-229C-21

Query Match: 67.9%; Score 45; DB 2; Length 321;
Best Local Similarity: 75.0%; Pred. No. 56;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KQGNFNW 8

DB 204 KQGVNFW 211

RESULT 11
US-08-659-235C-21
1 Sequence 21, Application US/08659235C
2 Patent No. 5877281
3 GENERAL INFORMATION:
4 APPLICANT: Quertemous, Thomas
5 APPLICANT: Hogan, Bridgid
6 APPLICANT: Snodgrass, H. Ralph
7 APPLICANT: Zupancic, Thomas J.
8 TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
9 NUMBER OF SEQUENCES: 29
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: Pennie & Edmonds LLP
12 STREET: 1155 Avenue of the Americas
13 CITY: New York
14 STATE: New York
15 COUNTRY: United States
16 ZIP: 10036-2711
17 COMPUTER READABLE FORM:
18 MEDIUM TYPE: Floppy disk
19 OPERATING SYSTEM: IBM PC compatible
20 SOFTWARE: Patent in Release #1.0, Version #1.30
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: US/08/659,235C
23 FILING DATE: 05 JUN-1996
24 CLASSIFICATION: 435
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Poissant, Brian M.
27 REGISTRATION NUMBER: 28,462
28 REFERENCE/POCKET NUMBER: 8907-0026-999
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: (212) 790-9090
31 TELEFAX: (212) 869-8854/9741
32 TELETYPE: 421792
33 INFORMATION FOR SEQ ID NO: 21:
34 SEQUENCE CHARACTERISTICS:
35 LENGTH: 321 amino acids
36 TYPE: amino acid
37 STRANDEDNESS:
38 TOPOLOGY: unknown
39 MOLECULE TYPE: protein
40 US-08-659-235C-21

Query Match: 67.9%; Score 36; DB 2; Length 321;
Best Local Similarity: 75.0%; Pred. No. 56;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KQGNFNW 8

```
DB 204 KQKVNAW 211
111 111
US-08-480-229C-10
: Sequence 10, Application US/08480229C
: Patent No. 5874562
: GENERAL INFORMATION:
: APPLICANT: Quertermous, Thomas
: APPLICANT: Hoan, Briqid
: APPLICANT: Snodgrass, H. Ralph
: APPLICANT: Zupancic, Thomas J.
: TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pennie & Edmonds LLP
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: United States
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/659,235C
: FILING DATE: 05-JUN-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Poissant, Brian M.
: REGISTRATION NUMBER: 28,462
: REFERENCE/DOC KEY NUMBER: 8907-0026-999
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864/9741
: TELEX: 66141 Pennie
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 480 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-480-229C-10

Query Match 67.9%; Score 36; ID 2; Length 480;
Best Local Similarity 75.0%; Pred No. 84;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KQGNFNAW 8
111 111
DB 460 KQKVNAW 367

RESULT 14
US-08-480-229C-14
: Sequence 14, Application US/08480229C
: Patent No. 5874562
: GENERAL INFORMATION:
: APPLICANT: Quertermous, Thomas
: APPLICANT: Hoan, Briqid
: APPLICANT: Snodgrass, H. Ralph
: APPLICANT: Zupancic, Thomas J.
: TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pennie & Edmonds LLP
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: United States
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/480,229C
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Poissant, Brian M.
: REGISTRATION NUMBER: 28,462
: REFERENCE/DOC KEY NUMBER: 8907-0026-999
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864/9741
: TELEX: 66141 Pennie
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 480 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-480-229C-10

Query Match 67.9%; Score 36; ID 2; Length 480;
Best Local Similarity 75.0%; Pred No. 84;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KQGNFNAW 8
111 111
DB 460 KQKVNAW 367

RESULT 13
US-08-659-245C-10
: Sequence 10, Application US/08659245C
: Patent No. 5877281
: GENERAL INFORMATION:
: APPLICANT: Quertermous, Thomas
: APPLICANT: Hoan, Briqid
: APPLICANT: Snodgrass, H. Ralph
: APPLICANT: Zupancic, Thomas J.
: TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pennie & Edmonds LLP
: STREET: 1155 Avenue of the Americas
```

TELEFAX: (212) 869 8864/9741
TELEX: 66141 Pennie
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 513 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-229C-14

Query Match 67.9%; Score 36; DB 2; Length 513;
Best Local Similarity 75.0%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KQGNFNAW 8
||| |||
Db 393 KQGNFNAW 400

RESULT 15

US-08-659-235C-14
Sequence 14, Application US/08659235C
Patent No. 5877281
GENERAL INFORMATION:
APPLICANT: Quentromous, Thomas
APPLICANT: Hogan, Bridget
APPLICANT: Snodgrass, H. Ralph
APPLICANT: Zupancic, Thomas J.
TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
TITLE OF INVENTION: CELL LOCUS-1
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/659,235C
FILING DATE: 05-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/BOOKET NUMBER: 8907-0034-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 Pennie
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 513 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-659-235C-14

Query Match 67.9%; Score 36; DB 2; Length 513;
Best Local Similarity 75.0%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KQGNFNAW 8
||| |||

Db 393 KQGNFNAW 400

Search completed: September 5, 2002, 15:27:44
Job time: 318 sec

WP1: 2500 205436,12.

Tumor associated antigen peptides, especially derived from uroplakin, useful as vaccines to prevent or cure cancers including breast, bladder, prostate, pancreas, ovary, thyroid, colon and stomach.

Claim 17; Page 99; 113pp; English.

Tumour associated antigen peptides (TAA) may be used for the treatment, prevention and cure of cancer or cancer metastases. The cancer may be breast, bladder, prostate, pancreas, ovary, thyroid, colon, stomach, head or neck cancer or a carcinoma. The tumour associated antigens are presentable to the immune system by HLA-A2 molecules and are generally at positions 8 to 10 amino acids in length. The amino acids located at positions 2 and 9 of the tumour associated antigens are the anchor residues which participate in the binding to MHC class I molecules, more specifically HLA-A2. More tumour associated antigens are described in GENESQ records AAY82806-Y82882. These tumour associated antigens described in records AAY82806-Y82824 and AAY82825-Y82864 are derived from Uroplakin, such as Uroplakin I, Uroplakin Ia, Uroplakin II and Uroplakin Ib. Those described in records AAY82825-Y82829 are derived from prostate specific antigen (PSA). Those described in records AAY82830-Y82835 are derived from prostate specific membrane antigen (PSMA). Those described in records Y82836-AAY82839 are derived from prostate acid phosphatase (PAP). Those described in records AAY82840-Y82845 are derived from lactadherin (LA-46). Those described in records AAY82847-Y82854 are derived from Mucin and those described in records AAY82857-Y82882 are derived from teratocarcinoma derived growth factor (CRPTO-1).

Sequence 9 AA:

Query Match: 100.0%; Score: 53; DP: 31; Length: 9;
 Best local Similarity: 100.0%; Pred. No: 6.4e+05;
 Matches: 9; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

QY	1	KUGNENAWV	9	
bb	1	kqgninawv	9	
RESULT 2				
		AAR77253		
Id		AAR77253 standard; Protein: 217 AA		
XX		AC		
XX		AAR77253;		
XX		21-NOV-1995 (first entry)		
XX		HMFG 46 kDa antigen C-terminal region.		
DE		XX		
XX		XX		
KW		HMFG; human milk fat globule; antigen; immunoglo; varicose; viracide;		
KW		epithelium; tumor; breast cancer; monoclonal antibody; MAb.		
XX		XX		
OS		Homo sapiens.		
XX		XX		
PN		W09515171-A.		
PD		08-JUN-1995.		
XX		XX		
PF		05-DEC-1994; 94WO-US13967.		
XX		XX		
PP		03-DEC-1994; 94US-0142432		
XX		XX		
PA		(CANC-) CANCER RES FUND CONTRA COSTA.		
XX		XX		
PI		Ceriani EL, Larocca DJ, Peterson JA;		
YX		XX		
DR		WPI; 1995-215151/22.		
XX		XX		
PT		46 kD apparent molecular weight human milk fat globule antigen		

PT used in assays to determine the presence of a cancerous tumour of
PT epithelial origin, and in a vaccine against neoplastic tumours
XX
XX
PS Claim 7; Page 41; 68pp; English.
XX
XX A partial cDNA clone B467 (AAQ61155) for the 46 kDa HMFG antigen,
CC a major component of the apical surface of the normal breast
CC epithelial cell, was obtained by PCR of cDNA from a lactating
CC breast cDNA library. The C-terminal region of the encoded protein
CC showed 43% identity to conserved regions of human Factor-V and 38%
CC to Factor-VIII.
XX
XX
XX Sequence 217 AA;
SU

```
Query Match      100.0%, Score 53;  pr 16;  length 217;
Fast local Similarity 100.0%;  pval 0.12;
Matches 9;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps
```

QY	1	KQCNFNNAV	9
Db	101	KQGNFNNAV	109

RESULT		3
	AAR77254	
ID	AAR77254	standard; Protein; 218 AA.
XX	XX	
AC	AAR77254;	
XX	XX	
DT	21-NOV-1995	(first entry)
XX	XX	
DE	HMFG 46 kDa antigen partial sequence.	

CC breast cDNA library.
XX Sequence 218 AA,
SQ

Query Match 100.0%; Score 53; DB 16; Length 218;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KOGNFNAWV 9
DB 102 kqgnfnawv 110

RESULT 4
AAR77252
ID AAR77252 standard; Protein: 387 AA.

XX AAR77252;

XX 21-NOV-1995 (first entry)

XX HMFG 46 kDa antigen.

XX HMFG; human milk fat globule; antigen; immunogen; vaccine; virucide;
XX epithelium; tumor; breast cancer; monoclonal antibody; MAb

XX Homo sapiens.

XX W09515171-A.

XX 08-JUN-1995.

XX 05-DEC-1994; G4W0-US13667

XX 03-DEC-1993; G4US-0152402.

XX (CANC-) CANCER RES FUND CONTRA COSTA.

XX Coriani PL, Larocca DJ, Peterson JA;

XX WPI; 1995-215151728

XX N-PSDB; AAG91198.

XX 46 kD apparent molecular weight human milk fat globule antigen
PT used in assays to determine the presence of a cancerous tumour of
PT epithelial origin, and in a vaccine against neoplastic tumours

XX Claim 6; Page 46-47; 62pp; English

XX A complete cDNA sequence for the 46 kDa HMFG antigen, a major
CC component of the apical surface of the normal breast epithelial
CC cell, was obtained by PCR and RACE methods. cDNA clones can be
CC used to prepare MABs for use in immunotherapy, immunohistopathology,
CC prognosis, diagnosis, imaging and therapy. Recombinant antigen can
CC be expressed in prokaryotic or (glycosylated) in eukaryotic cells.

XX Sequence 387 AA;

Query Match 100.0%; Score 53; DB 16; Length 387;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KOGNFNAWV 9
DB 271 kqgnfnawv 279

RESULT 5
AAY94453
ID AAY94453 standard; Protein: 387 AA.

XX

AC AAY94453;
XX 11 SEP 2000 (first entry)
XX Human lactadherin protein.
XX Human lactadherin; M3P-ER; anti-tumour; immune response;
KW exosome; dendritic cell.
XX Homo sapiens.

XX Key Location/Qualifiers
FT Peptide 1..23
FT Protein /label= Secretion_signal
FT /label= Lactadherin
FT Binding-site 46..48
FT /label= Integrin_binding_site

XX EP1004664-A1.

XX 31-MAY-2000.

XX 24-NOV-1998; 98EP-0402925.

XX 24-NOV-1998; 98EP-0402925.

XX (LIRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX (CURIE) INST CURIE.

XX WPI; 2000-352597/31.

XX N-PSDB; AAA27140.

XX Chimeric isolated (human) lactadherin polypeptide that functions as an
PT adaptor of cross-priming to eliminate pathogenic antigens, e.g. in
PT tumors -

XX Example 3; Page 12; 20pp; English.

XX Lactadherin protein was found in exosomes produced by dendritic cells.
CC The protein is involved in the phagocytosis of particulate antigens by
CC dendritic cells. Exosomes produced by dendritic cells exposed to
CC tumour antigens induce potent immune responses. Lactadherin or variants
CC of it may be used in the mediation of an immune response. Variants of
CC lactadherin may be used for inhibition and/or stimulation of the
CC cross priming of antigens and stimulation of the phagocytosis of
CC antigens by dendritic cells. Compositions derived from lactadherin can
CC also be used to monitor an immune response, more specifically a CTL
CC (cytotoxic T lymphocyte) response and also to produce CTLs specific
CC for a selected antigen. The present sequence is the human
CC lactadherin protein.

XX Sequence 387 AA;

Query Match 100.0%; Score 53; DB 21; Length 387;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KOGNFNAWV 9
DB 271 kqgnfnawv 279

RESULT 6
AAB71682
ID AAB71682 standard; protein: 363 AA.

XX AAB71682;

XX 10-MAY-2001 (first entry)

XX Bos taurus interleukin-2 receptor gamma.

XX Cytokine receptor common gamma chain like; CROCOL; human;
 KW tumours; infectious; inflammatory; immune disorder;
 KW neurodegenerative; cardiovascular; disorder.
 OS Bos taurus.
 XX WC200112672 A2.
 PH 22-FEB-2001.
 XX 17-AUG-2000; 2000WO-US22493.
 XX 18-AUG-1999; 990S-0376440.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Ruben SA, Rosen CA, Moore PA;
 PI WPI; 2001-147547/15.
 XX New nucleic acid molecule encoding a human cytokine receptor common
 PT gamma chain like polypeptide, useful for treating, preventing and/or
 PT diagnosing e.g. tumors, inflammatory diseases and immunodeficiency
 PT conditions -
 XX Disclosure; Fig 2; 288pp; English.
 XX The present invention relates to a human cytokine receptor
 CC common gamma chain like protein (CROCOL). The invention is useful
 CC for treating, preventing and/or diagnosing conditions such
 CC as tumors, infections, inflammatory diseases,
 CC immune disorders, neurodegenerative disorder and cardiovascular
 CC disorder.
 XX SQ Sequence 363 AA;

Query Match 73.6%; Score 39; DB 22; Length 363;
 Best Local Similarity 85.7%; Pred. No. 59;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Caps 0;
 QY 2 QGNENAW 8
 III:II
 DB 301 qnfsaw 307

RESULT 7
 AAR59094
 ID AAR59094 standard; Protein; 369 AA.
 AC AAR59094;
 XX 04-MAY-1995 (first entry)
 DT Murine IL-2R gamma.
 DE Murine IL-2R gamma; X-linked severe combined immunodeficiency;
 KW XSCID; interleukin.
 KW Mus musculus.
 OS Key location/Qualifiers
 PH Peptide 1..21
 FT /note= "signal peptide"
 FT Domain 258..284
 FT /note= "transmembrane domain"
 FT /note= "Corresponding codon CAG"
 FT Modified-site 71..73
 FT /label= N-glycosylation_site
 FT Modified-site 75..77
 FT /label= N-glycosylation_site

FT Modified-site 84..86
 FT /label= N-glycosylation_site
 FT Modified-site 96..98
 FT /label= N-glycosylation_site
 FT Modified-site 159..161
 FT /label= N-glycosylation_site
 FT Modified-site 255..257
 FT /label= N-glycosylation_site
 XX WC9420641-A.
 PN 15 SEP-1994.
 XX 10-MAR-1994; 94WO-US02891.
 XX 12-MAR-1993; 93US-0031143.
 XX 14-SEP-1993; 93US-0121435.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Leonard WJ, McBride WO, Nozuchi M;
 PI WPI; 1994-303046/37.
 DR N-PSOR; AA271977.
 XX Diagnosis of X-linked severe combined immunodeficiency (XSCID) -
 PT comprises detecting mutated IL-2R gamma gene, also vectors and
 PT transgenic animals containing the mutated gene
 XX Example 1; Fig 7; 98pp; English.
 CC AA271977 is the DNA sequence of murine IL-2R gamma AAR59094,
 CC this was used in the development of a claimed method for the
 CC diagnosis of X-linked severe combined immunodeficiency (XSCID),
 CC in female carriers and male sufferers.
 XX SQ Sequence 369 AA;

Query Match 73.6%; Score 39; DB 15; Length 369;
 Best Local Similarity 85.7%; Pred. No. 60;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Caps 0;

QY 2 QGNENAW 8
 III:II
 DB 304 qnfsaw 310
 RESULT 8
 ARG29550
 ID ARG29550 standard; Protein; 673 AA
 AC ARG29550;
 XX 18-FEB-2002 (first entry)
 DT Novel human diagnostic protein #29541.
 DE Human, chromosome mapping, gene mapping, gene therapy, forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 KW Homo sapiens.
 OS WC200175067 A2.
 PH 11-OCT-2001.
 XX 10-MAR-2000; 2001WO-US06431
 XX 31-MAR-2000; 2000US-0540217.
 XX 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.

XX
PI Drmanac RT, Liu C, Tang YI;
XX
DR WPI: 2001-639362/73;
DR N-PSDB: AAS93737.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20: SEQ ID No 5600q; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ARG00010-ARG0377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 673 AA;

Query Match 71.7%; Score 38; DB 22; Length 673;
Best local similarity 85.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 GNFAW 9
DB 432 gnfsaw 438
|||||

RESULT 9
ARG24996
ID ARG24996 standard; Protein: 1087 AA.

XX
AC ARG24996;

XX
DT 18-FEB-2002 (first entry)

XX
DE Novel human diagnostic protein #24987.

XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.

XX
OS Homo sapiens.

XX
PN W0200175067-A2.

XX
PD 11-OCT-2001.

XX
PF 30-MAR-2001; 2001WO-0508631.

XX
XX 31-MAR-2000; 2000US-0540217.

XX
XX 23-APR-2000; 2000US-0540177.

XX
PA (HYSEP-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YI;
XX
DR WPI: 2001-639362/73;
DR N-PSDB: AAS89183.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20: SEQ ID No 55355; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ARG00010-ARG0377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1087 AA;

Query Match 71.7%; Score 38; DB 22; Length 1087;
Best local similarity 85.7%; Pred. No. 2.8e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 GNFAW 9
DB 846 gnfsaw 852
|||||

RESULT 10
AAM65565
ID AAM65565 standard; Protein: 51 AA.

XX
AC AAM65565;

XX
DT 05-NOV-2001 (first entry)

XX
DE Human brain expressed single exon probe encoded protein SEQ ID No: 37670.

XX
XX Human: brain expressed exon; gene expression analysis; probe;

XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

XX epilepsy; cancer.

XX
OS Homo sapiens.

XX
PN W0200157275-A2.

XX
PD 09-AUG-2001.

XX
PF 30-JAN-2001; 2001WO-0500667.

XX
XX 04-FEB-2000; 2000US-0190312.

XX
XX 29-MAY-2000; 2000US-0207456.

XX
XX 30-JUN-2000; 2000US-0608408.

XX
XX 03-AUG-2000; 2000US-0632366.

XX
XX 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US 0234259.
 DR 04-OCT-2000; 2000JP-0324263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel BK, Chen W, Rank DK,
 XX WPI; 2001-483446/52.
 XX
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 PT
 XX
 XX Example 4. SEQ ID NO. 37670, 650pp + Sequence Listing; English.
 XX
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention
 XX
 XX Sequence 51 AA;

Query Match 67.9%; Score 36; DB 22; Length 51;
 Best Local Similarity 75.0%; Pred. No. 25;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KGNFNAW 8
 III III
 DB 48 kqkvvuaw 45

RESULT 11
 AAB65971
 ID AAB65971 standard; protein; 68 AA.
 AC AAB65971;
 XX
 XX 03-APR-2001 (first entry)
 XX
 XX Human secreted protein BLAST search protein SEQ ID NO. 111.
 DE
 XX Cytostatic; immunosuppressive; neotropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antileuc;
 KW vulvatory; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.
 XX
 XX Homo sapiens.
 OS
 XX W0200077024-A1.
 PN
 XX 21-DEC-2000.
 XX
 XX 01-JUN-2000; 2000WO-US14964.
 PF
 XX 11-JUN-1999; 99US-0138627.
 PR
 XX (HUMA-) HUMAN GENOME SCI INC
 PA
 XX Person GA, Paben SM, Komatsoulis GA;
 PI
 XX WPI; 2001-032408/04.
 DR
 XX Isolated nucleic acid molecule encoding a human secreted protein is
 PT used in preventing, treating or ameliorating a medical condition -
 PT
 XX Disclosure, Page 514, 558pp, English.
 PS
 XX The invention relates to the isolation of genes AAB65972-A45150 encoding
 CC the human secreted proteins AAB65920-A65968. This sequence represents a

CC peptide fragment homologous to the protein encoded by the gene given
 CC in the description line. The sequence is a search result from a BLASTX
 CC homology search. The genes and proteins are useful for preventing,
 CC ameliorating or treating medical conditions, e.g. by protein or gene
 CC therapy. The genes are isolated from a range of human tissues disclosed
 CC in the specification. The nucleic acids, proteins, antibodies and
 CC (antagonists are useful in the diagnosis, treatment and prevention of:
 CC (a) cancer, e.g. breast and ovarian cancer, and other cancers of the
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,
 CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; (c) cardiovascular disorders such
 CC as myocardial ischaemia; (d) wound healing; (e) neurological diseases
 CC e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such
 CC as viral, bacterial, fungal and parasitic infections.
 XX
 XX Sequence 68 AA;

Query Match 67.9%; Score 36; DB 22; Length 68;
 Best Local Similarity 55.6%; Pred. No. 43;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KGNFNAW 9
 : : : : :
 DB 10 eknnynaw 18

RESULT 12
 AAB04068
 ID AAB04068 standard; protein; 113 AA.
 XX
 AC AAB04068;
 XX
 XX 11-APR-2001 (first entry)
 DT
 XX Pat zeytor 10 cytokine receptor peptide fragment.
 DE
 XX
 KW zeytor 10 cytokine receptor; cytokine; receptor; antibody; ligand;
 KW binding; detection; modulation; recombinant cell;
 KW haematopoietic cell; lymphoid cell; myeloid cell; lymph;
 KW immune system; blood; bone; inflammatory response; inflammation;
 KW spleen; rat.
 XX
 XX Rattus rattus.
 OS
 XX W020008491-A1.
 PN
 XX 16-NOV-2000.
 PD
 XX
 XX 11-MAY-2000; 2000WO-US12924.
 PF
 XX 11-MAY-1999; 99US-0309861.
 PR
 XX (ZYMO) ZYMOGENETICS INC.
 PA
 XX Presnell SR, Foster DC, Hammond AK, Lok S;
 PI
 XX WPI; 2001-016096/02.
 DR
 XX N-PSDR; AAA54463.
 DP
 XX New cytokine receptor mouse zeytor 10, useful for detecting ligands
 PT that stimulate proliferation or development of haematopoietic,
 PT lymphoid and myeloid cells
 PT
 XX Example 3; Page 115; 134pp; English.
 PS
 XX Isolating a nucleotide which encodes the zeytor 10 cytokine
 CC receptor enables the production of recombinant cells expressing the
 CC receptor. Those cells can then be used to detect the presence of a
 CC modulator of zeytor10 protein by culturing the cells in the presence
 CC of a test ligand and comparing levels of activity of mouse zeytor10

CC in the presence and absence of the test sample. Similarity, detection
 CC of cytochrome receptor ligand within a test sample can be achieved.
 CC The method comprising contacting a test sample containing an amino
 CC acid sequence from 7815 to 8125 to a fragment of the cytochrome 12
 CC cytokine receptor and detecting the binding of the polypeptide to a
 CC ligand in the sample. Specified peptide fragments of the cytochrome 10
 CC cytokine receptor and the methods described are used to identify
 CC ligands that stimulate the proliferation and/or development of
 CC haematopoietic, lymphoid and myeloid cells. Peptide fragments of
 CC the cytokine receptor are useful for treating lymphoid, immune,
 CC inflammatory, splenic, blood or bone disorders and for generating
 CC antibodies directed against the receptor.
 XX
 SQ Sequence 113 AA;

Query Match 67.9%; Score 36; DB 22; Length 113;
 Best Local Similarity 71.4%; Pred. No. 57;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0.

QY 3 GNFNAW 9
 III III
 Db 41 gniqawi 47

RESULT 13
 AAB71691
 ID AAB71691 standard; peptide; 181 AA.

XX AAB71691;
 AC AAB71691;
 DT 10-MAY-2001 (first entry)
 XX
 DE CROGCL related peptide #4.

XX Cytokine receptor common gamma chain like; CROGCL; human;
 KW tumours; infections; inflammatory; immune disorder;
 KW neurodegenerative; cardiovascular; disorder.
 XX
 OS Homo sapiens.

XX WO200112672-A2
 PN 22-FEB-2001.
 XX
 PF 17-AUG-2000; 2000WO-0532493.
 XX
 PP 18-AUG-1999; 0995-0376430
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Puben SA, Posen CA, Moore PA;
 XX
 XX WPI: 2001 147647/15

XX New nucleic acid molecule encoding a human cytokine receptor common
 PT gamma chain like polypeptide, useful for treating, preventing and/or
 PT diagnosing of tumors, inflammatory diseases and immunodeficiency
 PT conditions.
 XX
 PS Disclosure; page 18; 288pp; English.

XX The present invention relates to a human cytokine receptor
 CC common gamma chain like protein (CROGCL). The invention is useful
 CC for treating, preventing and/or diagnosing conditions such
 CC as tumours, infections, inflammatory diseases,
 CC immune disorders, neurodegenerative disorder and cardiovascular
 CC disorder.
 XX
 SQ Sequence 181 AA;

Query Match 67.9%; Score 36; DB 22; Length 181;

Best Local Similarity 62.5%; Pred. No. 94;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 GNFNAW 9
 III III
 Db 91 gniqawi 98

RESULT 14
 AAW94697
 ID AAW94697 standard; Protein; 321 AA.
 XX
 AC AAW94697;
 XX
 DT 05-MAY-1999 (first entry)

XX Human milk fat globule protein MPG-E8.
 DE
 XX
 KW Del-1, developmentally regulated endothelial cell locus 1; cancer;
 KW discoidin 1; factor VIII-like domain; epidermal growth factor; EGF;
 KW diabetic retinopathy; rheumatoid arthritis; endometriosis;
 KW angiogenesis.
 XX
 OS Homo sapiens.

XX
 PN US5877281-A.
 XX
 PD 02-MAR-1999.

XX
 PF 05-JUN-1996; 9605-0659235.
 XX
 PP 05-JUN-1996; 9605-0659235.
 PP 07-JUN-1996; 9605-0450229.
 XX

PA (PROG-) PROGENITOR INC.
 PA (UYVA-) UNIV VANDERBILT.

PI Rejan B, Quattrone T, Shodgrass BE, Sepanick TJ;
 DE WPI: 1999-189720/16.

XX Del-1 polypeptide sequences useful for treatment of cancer,
 PT diabetic retinopathy, rheumatoid arthritis and endometriosis
 XX
 PS Example, Column 63 64, 73pp, English.

XX The present sequence represents human milk fat globule protein MPG-E8,
 CC which has the ability to the developmentally regulated endothelial cell
 CC locus 1 (del-1). The del-1 protein has epidermal growth factor (EGF) like
 CC domains and discoidin 1/factor VIII-like domains. The del-1 proteins
 CC have an inhibitory effect on angiogenesis (blood vessel growth), this
 CC activity may be useful clinically to prevent neovascularisation of
 CC tissues such as tumour nodules and prevention of metastases. The anti-
 CC angiogenic activity of del-1 may be used to treat abnormal conditions
 CC that result from angiogenesis, including cancer, diabetic retinopathy,
 CC rheumatoid arthritis and endometriosis. Since del-1 promotes
 CC angiogenesis it can be used to treat cardiac ischemia, thrombotic
 CC stroke, wound healing and peripheral vascular disease. Del-1 is also
 CC useful for promoting bone formation. Del-1 binds to alpha V beta 3
 CC integrin, and is an apoptosis inducer.
 XX
 SQ Sequence 321 AA;

Query Match 67.9%; Score 36; DB 20; Length 321;
 Best Local Similarity 75.0%; Pred. No. 1176-02;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KGNFNAW 8
 III III
 Db 204 kqkvnaw 211

```

RESULT 15
AA04066
ID AA04066 standard; Protein; 357 AA.
XX
AC AA04066;
XX
DT 11-APR-2001 (first entry)
XX
DE Zcytor 10 cytokine receptor.
XX
FW Zcytor 10 cytokine receptor, cytokine, receptor, antibody; ligand;
KW binding; detection; modulation; recombinant cell;
KW haematopoietic cell; lymphoid cell; myeloid cell; lymph;
KW immune system; blood; bone; inflammatory response; inflammation;
KW spleen; human.
XX
OS Homo sapiens.
XX
PN W0200068481-A1.
XX
PD 16-NOV-2000.
XX
PF 11-MAY-2000; 2000W0-US12924.
XX
PR 11-MAY-1999; 590S-0309861.
XX
PA (ZYMO ) ZYMOGENETICS INC.
XX
PI Fresnell SK, Foster DC, Hammond AK, Lok S;
DR WPI; 2001-016096/02.
DR N-PSDB; AAA54451.
XX
PI New cytokine receptor mouse zcytor 10, useful for detecting ligands
PI that stimulate proliferation or development of haematopoietic,
PI lymphoid and myeloid cells
XX
PS Claim 1; Page 109-110; 134pp; English.
XX
CC Isolating a nucleotide which encodes the zcytor 10 cytokine
CC receptor enables the production of recombinant cells expressing the
CC modulator of zcytor10 protein by culturing the cells in the presence
CC of a test ligand and comparing levels of activity of mouse zcytor10
CC in the presence and absence of the test sample. Similarly, detection
CC of zcytor10 receptor ligand within a test sample can be achieved.
CC The method comprising contacting a test sample containing an amino
CC acid sequence from 6315 to 6330 of the zcytor 10
CC cytokine receptor and detecting the binding of the polypeptide to a
CC ligand in the sample. Specified peptide fragments of the zcytor 10
CC cytokine receptor and the methods described are used to identify
CC ligands that stimulate the proliferation and/or development of
CC haematopoietic, lymphoid and myeloid cells. Peptide fragments of
CC the cytokine receptor are useful for treating lymphoid, immune,
CC inflammatory, splenic, blood or bone disorders and for generating
CC antibodies directed against the receptor.
XX
SQ Sequence 357 AA;

```

```

Query Match 67.9%; Score 36; DB 22; Length 357;
Best Local Similarity 71.4%; Pred. No. 1.9e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 3 GNFNAWV 9
    III II
Db 280 gnfqawi 286

```

Search completed: September 5, 2002, 15:26:28
Job time: 368 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Computer Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:22:55 : Search time 04.74 seconds
(without alignments)
12.393 Million cell updates/sec

Title: US-09-744-804-35
Perfect score: 53
Sequence: 1 KOGHNAWV 9

Scoring table: PROSUMC2
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	100.0	218	A47285	milk fat globule p
2	42	79.2	734	S00704	photosystem 1 prot
3	42	79.2	734	T07544	photosystem 1 prot
4	40	75.5	268	T09026	hypothetical prote
5	40	75.5	386	H85684	unknown protein on
6	34	73.6	129	AP0154	probable bacteriop
7	39	73.6	369	T49280	interleukin 2 rece
8	39	73.6	734	A2NT97	photosystem 1 p700
9	39	73.6	734	S00445	photosystem 1 prot
10	38	71.7	61	S60184	photosystem 1 prot
11	38	71.7	64	C90821	hypothetical prote
12	38	71.7	64	D90832	hypothetical prote
13	38	71.7	64	H85689	unknown protein on
14	38	71.7	74	E90901	hypothetical prote
15	38	71.7	74	H85743	unknown protein on
16	38	71.7	401	S65138	glycoprotein anti
17	38	71.7	409	T11743	p647 protein - pig
18	38	71.7	427	S74211	PAS-6/7 protein p
19	38	71.7	427	A2R217	photosystem 1 p700
20	38	71.7	734	A21V67	photosystem 1 p700
21	38	71.7	735	S78551	photosystem 1 prot
22	37	69.8	373	A10440	probable aliphatic
23	37	69.8	503	S64860	hypothetical prote
24	37	69.8	733	S78476	photosystem 1 p700
25	27	63.8	734	T67580	photosystem 1 p700
26	37	69.8	735	S41481	P700 chlorophyll a
27	37	69.8	736	B28341	photosystem 1 p700
28	37	69.8	737	T06839	probable photosyst
29	37	69.8	741	S20423	photosystem 1 prot

30	37	69.8	817	2	S77106	hypothetical prote
31	36	67.9	278	2	A60195	transforming prote
32	36	67.9	354	2	T07080	cyclic line receptor
33	36	67.9	434	2	AG0071	Exl1 transport pro
34	36	67.9	472	1	H65098	hexanate transpo
35	36	67.9	472	2	C91125	transport protein
36	36	67.9	472	2	P85970	transport of hexur
37	36	67.9	503	2	S13012	aromatase (EC 1.14
38	36	67.9	508	1	A36121	hypothetical prote
39	36	67.9	526	2	C83790	photosystem 1 prot
40	36	67.9	731	2	S18243	photosystem 1 prot
41	36	67.9	733	2	S06902	photosystem 1 prot
42	36	67.9	734	1	S26072	photosystem 1 prot
43	36	67.9	734	2	S73206	photosystem 1 prot
44	36	67.9	741	2	T39616	photosystem 1 prot
45	36	67.9	741	2	A02450	photosystem 1 core

ALIGNMENTS

RESULT 1

A47285
milk fat globule protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 31-Jan-1993 #sequence_revision: 11-Nov-1993 #text_change: 29-May-1999
C:Accession: A47285
R:Larocca, D.; Peterson, J.A.; Urrea, R.; Kuniyoshi, J.; Ristrain, A.M.; Ceriani, R.I.
Cancer Res. 51, 4994-4998, 1991
A:Title: A Mr 45,000 human milk fat globule protein that is highly expressed in huma
A:Reference number: A47285; MUID:91371351
A:Accession: A47285
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-218 <LAK>
A:Cross reference: de Smet, N.H. 45,000, PUD-AAB19771.1, PUD-A233397
C:Superfamily: milk fat globule protein; discoidin 1 amino-terminal homology; PGF hom
F:1-56/domain; discoidin 1 amino terminal homology (fragment) <DRS>
F:60-218/domain; discoidin 1 amino terminal homology <DRS>

Query Match 100.0%, Score 53, DP 2, Length 218;
Best local Similarity 100.0%, Pred. No. 0.019;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Q: 1 KOGHNAWV 9
B: 102 KOGHNAWV 110
RESULT 2
S00704
Photosystem 1 protein, A2, garden pea chloroplast
C:Species: chloroplast Pisum sativum (garden pea)
C:Date: 28-Apr-1989 #sequence_revision: 28-Aug-1989 #text_change: 26-Apr-1989
C:Accession: S00704; R30830
F:Lehman, J.; Fasman, G.F.; Beck-Jones, G.B.; Tolson, R.E.; Stummann, B.M.; Hennin
Plant Mol. Biol. 7, 3-10, 1986
A:Title: Sequence of two genes in pea chloroplast DNA coding for 64 and 82 kD polypep
A:Reference number: S00703
A:Accession: S00704
A:Molecule type: DNA
A:Residues: 1-734 <LEH>
A:Cross reference: EMEL-N95423, KID-A12134, PUD-AAC29204.1, PUD-A12136
A:Note: the authors translated the codon AAT for residue 115 as Ile and ATT for resid
C:Genetics:
A:Gene: psaA2
A:Gene: chloroplast
C:Superfamily: Photosystem 1 P700 ap protein
C:Keywords: chloroplast; electron transfer; membrane-associated complex; photosynthes

Query Match 79.2%, Score 42, DP 2, Length 734;

```

Best Local Similarity 87.5%; Score 40; P: 2; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 2 QGNFNW 9
    |||||
Db 71 QGNFNW 78

RESULT 3
T07544
photocycle 1 protein A2; cytoplasmic flavin chlorophyll
C:Species: chloroplast; photosynthesis; chlorophyll; flavin
C>Date: 14-May-1999; #sequence_revision 14 May 1999; #text_change 18-Aug-2000
C:Accession: T07544
R:Wakasugi, T.; Tsudzuki, J.; Ito, S.; Nakashima, K.; Tsudzuki, T.; Sugita, M.
Proc. Natl. Acad. Sci. U.S.A. 91, 9794-9798, 1994
A:Title: Loss of all indh genes as determined by sequencing the entire chloroplast genome
A:Reference number: 216030; MIM:05054047
A:Accession: T07544
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-734 <WAK>
A:Cross-references: EMBL:D17510; NID:q529643; PIDN:DAA04420.1; PID:q1262705
C:Genetics:
A:Gene: psbA
A:Superfamily: chloroplast
C:Keywords: chloroplast, membrane associated complex, photosystem I

Query Match 79.2%; Score 42; P: 2; Length 734;
Best Local Similarity 87.5%; Pred. No. 6.9;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGNFNW 9
    |||||
Db 71 QGNFNW 78

RESULT 4
H90826
hypothetical protein ECs1584 [imported] - Escherichia coli (strain O157:H7, substrain P2)
C:Species: Escherichia coli
C>Date: 18-Jul-2001; #sequence_revision 18-Jul-2001; #text_change 18-Jul-2001
C:Accession: H90826
R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuwara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome
A:Reference number: 93429; MIM:2116231; PMID:11258796
A:Accession: H90826
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-268 <HAY>
A:Cross-references: GB:HA000007; PIDN:PAR35007.1; PID:q13361048; GSPIR-GN00154
A:Experimental source: strain O157:H7, substrain KMD 0509952
C:Genetics:
A:Gene: ECs1584

Query Match 75.5%; Score 40; P: 2; Length 268;
Best Local Similarity 87.5%; Pred. No. 5.9;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGNFNW 9
    |||||
Db 145 QGNFNW 142

RESULT 5
H85684
unknown protein encoded by prophage CP 9346 [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli

```

```

C>Date: 16-Feb-2001; #sequence_revision 16-Feb-2001; #text_change 14-Sep-2001
C:Accession: H85684
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.D.; May
Miller, L.; Grotbeck, E.; Davis, N.W.; Lim, A.; Dimalanta, F.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MIM:21074935; PMID:11206551
A:Accession: H85684
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-326 <STO>
A:Cross-references: GB:AF005174; NID:q12514762; PIDN:AA056440.1; GSPIR-GN00145; OM01;
A:Experimental source: strain O157:H7, substrain FBI933
C:Genetics:
A:Gene: Z1843

Query Match 75.5%; Score 40; P: 2; Length 386;
Best Local Similarity 87.5%; Pred. No. 8.5;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGNFNW 9
    |||||
Db 253 QGNFNW 260

RESULT 6
AB0154
probable bacteriophage protein (partial) YB01259 [imported] - Yersinia pestis (strain
C:Species: Yersinia pestis
C>Date: 02-Nov-2001; #sequence_revision 02-Nov-2001; #text_change 02-Nov-2001
C:Accession: AB0154
R:Parkhill, T.; Wren, B.W.; Thomson, N.P.; Titball, R.W.; Holden, M.T.G.; Prentice, M.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, K.M.; Davis, P.; Dougan, G.
II, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AP0001; MIM:21470413; PMID:11586360
A:Accession: AB0154
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-127 <KUP>
A:Cross-references: GB:AL595842; PIDN:ZAN03093.1; PID:q1567933; GSPIR-GN00175
C:Genetics:
A:Gene: YP01259

```

```

Query Match 73.6%; Score 39; P: 2; Length 127;
Best Local Similarity 75.0%; Pred. No. 4.3;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGNFNW 9
    |||||
Db 19 QGNFNW 26

RESULT 7
I49280
interleukin-2 receptor gamma chain precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 27-Feb-1997; #sequence_revision 27-Feb-1997; #text_change 20-Feb-2000
C:Accession: I49280; A47534; UN0592; UN0775; 837582; 153398
R:Cao, X.; Kozak, C.A.; Liu, Y.
Proc. Natl. Acad. Sci. U.S.A. 90, 8464-8469, 1993
A:Title: Characterization of cDNAs encoding the murine interleukin-2 receptor (IL-2R)
A:Reference number: A47514; MIM:93491374
A:Accession: I49280
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-369 <CAO>
A:Cross-references: EMBL:U01795; NID:q272749; PIDN:AAA64279.1; PID:q27450
A:Accession: A47514
A:Status: translated from GB/EMBL/DBJ

```

A:Molecule type: mRNA
A:Residues: 1-369 <RES>
A:Cross-references: GB:120048, MID:9404067, PIRN:AAA34286.1, FID:9404068
F:Kamaki, S.; Kondo, M.; Takeuchi, T.; Asano, H.; Nakamura, M.; Suganuma, K.
Biochem. Biophys. Res. Commun. 193, 356-363, 1993
A:Title: Cloning of the mouse interleukin 2 receptor gamma chain: demonstration of function
A:Reference number: 38592, MUIP:94277575
A:Accession: JN0592
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-369 <KUM>
A:Cross-references: J04113365, NID:9403084, FIDN:AAA32700.1, FID:9403085
R:Kobayashi, N.; Nakagawa, S.; Minami, Y.; Taniguchi, T.; Kono, T.
Gene 130, 303-304, 1993
A:Title: Cloning and sequencing of the cDNA encoding a mouse IL-2 receptor gamma.
A:Reference number: JN0775, MUIP:94366191
A:Accession: JN0775
A:Molecule type: mRNA
A:Residues: 1-369 <KOR>
A:Cross-references: GB:113821, MID:9436945, FIDN:AAA02974.1, FID:9436946
R:Chiu, R.K.; Dougherty, G.J.
submitted to the EMBL Data Library, October 1993
A:Description: Regulation of CD44-mediated cellular adhesion by the IL-2 R gamma chain.
A:Reference number: S37582
A:Accession: S37582
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-350, 5'-352-366, 5'-368-369 <CHI>
A:Cross-references: EMBL:X75337
R:Disanto, J.P.; Certain, S.; Wilson, A.; Macdonald, H.P.; Avner, P.; Fischer, A.; de Sa
Eur. J. Immunol. 24, 3014-3018, 1994
A:Title: The murine interleukin 2 receptor gamma chain gene: organization, chromosomal location
A:Reference number: 154308, MUIP:95104285
A:Accession: I53398
A:Status: preliminary, translated from GB, EMBL, EMBJ
A:Molecule type: DNA
A:Residues: 1-369 <RES>
A:Cross-references: GB:S75852, NID:9861564, FIDN:AAA42404.1, FID:9861555
C:Genetics:
A:Gene: IL-2/gamma
A:Comments: 36, 1-369, 5'-152-156, 5'-158-159, 5'-160-161, 5'-162-163, 5'-164-165, 5'-166-167, 5'-168-169, 5'-170-171, 5'-172-173, 5'-174-175, 5'-176-177, 5'-178-179, 5'-180-181, 5'-182-183, 5'-184-185, 5'-186-187, 5'-188-189, 5'-190-191, 5'-192-193, 5'-194-195, 5'-196-197, 5'-198-199, 5'-200-201, 5'-202-203, 5'-204-205, 5'-206-207, 5'-208-209, 5'-210-211, 5'-212-213, 5'-214-215, 5'-216-217, 5'-218-219, 5'-220-221, 5'-222-223, 5'-224-225, 5'-226-227, 5'-228-229, 5'-230-231, 5'-232-233, 5'-234-235, 5'-236-237, 5'-238-239, 5'-240-241, 5'-242-243, 5'-244-245, 5'-246-247, 5'-248-249, 5'-250-251, 5'-252-253, 5'-254-255, 5'-256-257, 5'-258-259, 5'-260-261, 5'-262-263, 5'-264-265, 5'-266-267, 5'-268-269, 5'-270-271, 5'-272-273, 5'-274-275, 5'-276-277, 5'-278-279, 5'-280-281, 5'-282-283, 5'-284-285, 5'-286-287, 5'-288-289, 5'-290-291, 5'-292-293, 5'-294-295, 5'-296-297, 5'-298-299, 5'-300-301, 5'-302-303, 5'-304-305, 5'-306-307, 5'-308-309, 5'-310-311, 5'-312-313, 5'-314-315, 5'-316-317, 5'-318-319, 5'-320-321, 5'-322-323, 5'-324-325, 5'-326-327, 5'-328-329, 5'-330-331, 5'-332-333, 5'-334-335, 5'-336-337, 5'-338-339, 5'-340-341, 5'-342-343, 5'-344-345, 5'-346-347, 5'-348-349, 5'-350-351, 5'-352-353, 5'-354-355, 5'-356-357, 5'-358-359, 5'-360-361, 5'-362-363, 5'-364-365, 5'-366-367, 5'-368-369, 5'-370-371, 5'-372-373, 5'-374-375, 5'-376-377, 5'-378-379, 5'-380-381, 5'-382-383, 5'-384-385, 5'-386-387, 5'-388-389, 5'-390-391, 5'-392-393, 5'-394-395, 5'-396-397, 5'-398-399, 5'-400-401, 5'-402-403, 5'-404-405, 5'-406-407, 5'-408-409, 5'-410-411, 5'-412-413, 5'-414-415, 5'-416-417, 5'-418-419, 5'-420-421, 5'-422-423, 5'-424-425, 5'-426-427, 5'-428-429, 5'-430-431, 5'-432-433, 5'-434-435, 5'-436-437, 5'-438-439, 5'-440-441, 5'-442-443, 5'-444-445, 5'-446-447, 5'-448-449, 5'-450-451, 5'-452-453, 5'-454-455, 5'-456-457, 5'-458-459, 5'-460-461, 5'-462-463, 5'-464-465, 5'-466-467, 5'-468-469, 5'-470-471, 5'-472-473, 5'-474-475, 5'-476-477, 5'-478-479, 5'-480-481, 5'-482-483, 5'-484-485, 5'-486-487, 5'-488-489, 5'-490-491, 5'-492-493, 5'-494-495, 5'-496-497, 5'-498-499, 5'-500-501, 5'-502-503, 5'-504-505, 5'-506-507, 5'-508-509, 5'-510-511, 5'-512-513, 5'-514-515, 5'-516-517, 5'-518-519, 5'-520-521, 5'-522-523, 5'-524-525, 5'-526-527, 5'-528-529, 5'-530-531, 5'-532-533, 5'-534-535, 5'-536-537, 5'-538-539, 5'-540-541, 5'-542-543, 5'-544-545, 5'-546-547, 5'-548-549, 5'-550-551, 5'-552-553, 5'-554-555, 5'-556-557, 5'-558-559, 5'-560-561, 5'-562-563, 5'-564-565, 5'-566-567, 5'-568-569, 5'-570-571, 5'-572-573, 5'-574-575, 5'-576-577, 5'-578-579, 5'-580-581, 5'-582-583, 5'-584-585, 5'-586-587, 5'-588-589, 5'-590-591, 5'-592-593, 5'-594-595, 5'-596-597, 5'-598-599, 5'-600-601, 5'-602-603, 5'-604-605, 5'-606-607, 5'-608-609, 5'-610-611, 5'-612-613, 5'-614-615, 5'-616-617, 5'-618-619, 5'-620-621, 5'-622-623, 5'-624-625, 5'-626-627, 5'-628-629, 5'-630-631, 5'-632-633, 5'-634-635, 5'-636-637, 5'-638-639, 5'-640-641, 5'-642-643, 5'-644-645, 5'-646-647, 5'-648-649, 5'-650-651, 5'-652-653, 5'-654-655, 5'-656-657, 5'-658-659, 5'-660-661, 5'-662-663, 5'-664-665, 5'-666-667, 5'-668-669, 5'-670-671, 5'-672-673, 5'-674-675, 5'-676-677, 5'-678-679, 5'-680-681, 5'-682-683, 5'-684-685, 5'-686-687, 5'-688-689, 5'-690-691, 5'-692-693, 5'-694-695, 5'-696-697, 5'-698-699, 5'-700-701, 5'-702-703, 5'-704-705, 5'-706-707, 5'-708-709, 5'-710-711, 5'-712-713, 5'-714-715, 5'-716-717, 5'-718-719, 5'-720-721, 5'-722-723, 5'-724-725, 5'-726-727, 5'-728-729, 5'-730-731, 5'-732-733, 5'-734-735, 5'-736-737, 5'-738-739, 5'-740-741, 5'-742-743, 5'-744-745, 5'-746-747, 5'-748-749, 5'-750-751, 5'-752-753, 5'-754-755, 5'-756-757, 5'-758-759, 5'-760-761, 5'-762-763, 5'-764-765, 5'-766-767, 5'-768-769, 5'-770-771, 5'-772-773, 5'-774-775, 5'-776-777, 5'-778-779, 5'-780-781, 5'-782-783, 5'-784-785, 5'-786-787, 5'-788-789, 5'-790-791, 5'-792-793, 5'-794-795, 5'-796-797, 5'-798-799, 5'-800-801, 5'-802-803, 5'-804-805, 5'-806-807, 5'-808-809, 5'-810-811, 5'-812-813, 5'-814-815, 5'-816-817, 5'-818-819, 5'-820-821, 5'-822-823, 5'-824-825, 5'-826-827, 5'-828-829, 5'-830-831, 5'-832-833, 5'-834-835, 5'-836-837, 5'-838-839, 5'-840-841, 5'-842-843, 5'-844-845, 5'-846-847, 5'-848-849, 5'-850-851, 5'-852-853, 5'-854-855, 5'-856-857, 5'-858-859, 5'-860-861, 5'-862-863, 5'-864-865, 5'-866-867, 5'-868-869, 5'-870-871, 5'-872-873, 5'-874-875, 5'-876-877, 5'-878-879, 5'-880-881, 5'-882-883, 5'-884-885, 5'-886-887, 5'-888-889, 5'-890-891, 5'-892-893, 5'-894-895, 5'-896-897, 5'-898-899, 5'-900-901, 5'-902-903, 5'-904-905, 5'-906-907, 5'-908-909, 5'-910-911, 5'-912-913, 5'-914-915, 5'-916-917, 5'-918-919, 5'-920-921, 5'-922-923, 5'-924-925, 5'-926-927, 5'-928-929, 5'-930-931, 5'-932-933, 5'-934-935, 5'-936-937, 5'-938-939, 5'-940-941, 5'-942-943, 5'-944-945, 5'-946-947, 5'-948-949, 5'-950-951, 5'-952-953, 5'-954-955, 5'-956-957, 5'-958-959, 5'-960-961, 5'-962-963, 5'-964-965, 5'-966-967, 5'-968-969, 5'-970-971, 5'-972-973, 5'-974-975, 5'-976-977, 5'-978-979, 5'-980-981, 5'-982-983, 5'-984-985, 5'-986-987, 5'-988-989, 5'-990-991, 5'-992-993, 5'-994-995, 5'-996-997, 5'-998-999, 5'-1000-1001, 5'-1002-1003, 5'-1004-1005, 5'-1006-1007, 5'-1008-1009, 5'-1010-1011, 5'-1012-1013, 5'-1014-1015, 5'-1016-1017, 5'-1018-1019, 5'-1020-1021, 5'-1022-1023, 5'-1024-1025, 5'-1026-1027, 5'-1028-1029, 5'-1030-1031, 5'-1032-1033, 5'-1034-1035, 5'-1036-1037, 5'-1038-1039, 5'-1040-1041, 5'-1042-1043, 5'-1044-1045, 5'-1046-1047, 5'-1048-1049, 5'-1050-1051, 5'-1052-1053, 5'-1054-1055, 5'-1056-1057, 5'-1058-1059, 5'-1060-1061, 5'-1062-1063, 5'-1064-1065, 5'-1066-1067, 5'-1068-1069, 5'-1070-1071, 5'-1072-1073, 5'-1074-1075, 5'-1076-1077, 5'-1078-1079, 5'-1080-1081, 5'-1082-1083, 5'-1084-1085, 5'-1086-1087, 5'-1088-1089, 5'-1090-1091, 5'-1092-1093, 5'-1094-1095, 5'-1096-1097, 5'-1098-1099, 5'-1100-1101, 5'-1102-1103, 5'-1104-1105, 5'-1106-1107, 5'-1108-1109, 5'-1110-1111, 5'-1112-1113, 5'-1114-1115, 5'-1116-1117, 5'-1118-1119, 5'-1120-1121, 5'-1122-1123, 5'-1124-1125, 5'-1126-1127, 5'-1128-1129, 5'-1130-1131, 5'-1132-1133, 5'-1134-1135, 5'-1136-1137, 5'-1138-1139, 5'-1140-1141, 5'-1142-1143, 5'-1144-1145, 5'-1146-1147, 5'-1148-1149, 5'-1150-1151, 5'-1152-1153, 5'-1154-1155, 5'-1156-1157, 5'-1158-1159, 5'-1160-1161, 5'-1162-1163, 5'-1164-1165, 5'-1166-1167, 5'-1168-1169, 5'-1170-1171, 5'-1172-1173, 5'-1174-1175, 5'-1176-1177, 5'-1178-1179, 5'-1180-1181, 5'-1182-1183, 5'-1184-1185, 5'-1186-1187, 5'-1188-1189, 5'-1190-1191, 5'-1192-1193, 5'-1194-1195, 5'-1196-1197, 5'-1198-1199, 5'-1200-1201, 5'-1202-1203, 5'-1204-1205, 5'-1206-1207, 5'-1208-1209, 5'-1210-1211, 5'-1212-1213, 5'-1214-1215, 5'-1216-1217, 5'-1218-1219, 5'-1220-1221, 5'-1222-1223, 5'-1224-1225, 5'-1226-1227, 5'-1228-1229, 5'-1230-1231, 5'-1232-1233, 5'-1234-1235, 5'-1236-1237, 5'-1238-1239, 5'-1240-1241, 5'-1242-1243, 5'-1244-1245, 5'-1246-1247, 5'-1248-1249, 5'-1250-1251, 5'-1252-1253, 5'-1254-1255, 5'-1256-1257, 5'-1258-1259, 5'-1260-1261, 5'-1262-1263, 5'-1264-1265, 5'-1266-1267, 5'-1268-1269, 5'-1270-1271, 5'-1272-1273, 5'-1274-1275, 5'-1276-1277, 5'-1278-1279, 5'-1280-1281, 5'-1282-1283, 5'-1284-1285, 5'-1286-1287, 5'-1288-1289, 5'-1290-1291, 5'-1292-1293, 5'-1294-1295, 5'-1296-1297, 5'-1298-1299, 5'-1300-1301, 5'-1302-1303, 5'-1304-1305, 5'-1306-1307, 5'-1308-1309, 5'-1310-1311, 5'-1312-1313, 5'-1314-1315, 5'-1316-1317, 5'-1318-1319, 5'-1320-1321, 5'-1322-1323, 5'-1324-1325, 5'-1326-1327, 5'-1328-1329, 5'-1330-1331, 5'-1332-1333, 5'-1334-1335, 5'-1336-1337, 5'-1338-1339, 5'-1340-1341, 5'-1342-1343, 5'-1344-1345, 5'-1346-1347, 5'-1348-1349, 5'-1350-1351, 5'-1352-1353, 5'-1354-1355, 5'-1356-1357, 5'-1358-1359, 5'-1360-1361, 5'-1362-1363, 5'-1364-1365, 5'-1366-1367, 5'-1368-1369, 5'-1370-1371, 5'-1372-1373, 5'-1374-1375, 5'-1376-1377, 5'-1378-1379, 5'-1380-1381, 5'-1382-1383, 5'-1384-1385, 5'-1386-1387, 5'-1388-1389, 5'-1390-1391, 5'-1392-1393, 5'-1394-1395, 5'-1396-1397, 5'-1398-1399, 5'-1400-1401, 5'-1402-1403, 5'-1404-1405, 5'-1406-1407, 5'-1408-1409, 5'-1410-1411, 5'-1412-1413, 5'-1414-1415, 5'-1416-1417, 5'-1418-1419, 5'-1420-1421, 5'-1422-1423, 5'-1424-1425, 5'-1426-1427, 5'-1428-1429, 5'-1430-1431, 5'-1432-1433, 5'-1434-1435, 5'-1436-1437, 5'-1438-1439, 5'-1440-1441, 5'-1442-1443, 5'-1444-1445, 5'-1446-1447, 5'-1448-1449, 5'-1450-1451, 5'-1452-1453, 5'-1454-1455, 5'-1456-1457, 5'-1458-1459, 5'-1460-1461, 5'-1462-1463, 5'-1464-1465, 5'-1466-1467, 5'-1468-1469, 5'-1470-1471, 5'-1472-1473, 5'-1474-1475, 5'-1476-1477, 5'-1478-1479, 5'-1480-1481, 5'-1482-1483, 5'-1484-1485, 5'-1486-1487, 5'-1488-1489, 5'-1490-1491, 5'-1492-1493, 5'-1494-1495, 5'-1496-1497, 5'-1498-1499, 5'-1500-1501, 5'-1502-1503, 5'-1504-1505, 5'-1506-1507, 5'-1508-1509, 5'-1510-1511, 5'-1512-1513, 5'-1514-1515, 5'-1516-1517, 5'-1518-1519, 5'-1520-1521, 5'-1522-1523, 5'-1524-1525, 5'-1526-1527, 5'-1528-1529, 5'-1530-1531, 5'-1532-1533, 5'-1534-1535, 5'-1536-1537, 5'-1538-1539, 5'-1540-1541, 5'-1542-1543, 5'-1544-1545, 5'-1546-1547, 5'-1548-1549, 5'-1550-1551, 5'-1552-1553, 5'-1554-1555, 5'-1556-1557, 5'-1558-1559, 5'-1560-1561, 5'-1562-1563, 5'-1564-1565, 5'-1566-1567, 5'-1568-1569, 5'-1570-1571, 5'-1572-1573, 5'-1574-1575, 5'-1576-1577, 5'-1578-1579, 5'-1580-1581, 5'-1582-1583, 5'-1584-1585, 5'-1586-1587, 5'-1588-1589, 5'-1590-1591, 5'-1592-1593, 5'-1594-1595, 5'-1596-1597, 5'-1598-1599, 5'-1600-1601, 5'-1602-1603, 5'-1604-1605, 5'-1606-1607, 5'-1608-1609, 5'-1610-1611, 5'-1612-1613, 5'-1614-1615, 5'-1616-1617, 5'-1618-1619, 5'-1620-1621, 5'-1622-1623, 5'-1624-1625, 5'-1626-1627, 5'-1628-1629, 5'-1630-1631, 5'-1632-1633, 5'-1634-1635, 5'-1636-1637, 5'-1638-1639, 5'-1640-1641, 5'-1642-1643, 5'-1644-1645, 5'-1646-1647, 5'-1648-1649, 5'-1650-1651, 5'-1652-1653, 5'-1654-1655, 5'-1656-1657, 5'-1658-1659, 5'-1660-1661, 5'-1662-1663, 5'-1664-1665, 5'-1666-1667, 5'-1668-1669, 5'-1670-1671, 5'-1672-1673, 5'-1674-1675, 5'-1676-1677, 5'-1678-1679, 5'-1680-1681, 5'-1682-1683, 5'-1684-1685, 5'-1686-1687, 5'-1688-1689, 5'-1690-1691, 5'-1692-1693, 5'-1694-1695, 5'-1696-1697, 5'-1698-1699, 5'-1700-1701, 5'-1702-1703, 5'-1704-1705, 5'-1706-1707, 5'-1708-1709, 5'-1710-1711, 5'-1712-1713, 5'-1714-1715, 5'-1716-1717, 5'-1718-1719, 5'-1720-1721, 5'-1722-1723, 5'-1724-1725, 5'-1726-1727, 5'-1728-1729, 5'-1730-1731, 5'-1732-1733, 5'-1734-1735, 5'-1736-1737, 5'-1738-1739, 5'-1740-1741, 5'-1742-1743, 5'-1744-1745, 5'-1746-1747, 5'-1748-1749, 5'-1750-1751, 5'-1752-1753, 5'-1754-1755, 5'-1756-1757, 5'-1758-1759, 5'-1760-1761, 5'-1762-1763, 5'-1764-1765, 5'-1766-1767, 5'-1768-1769, 5'-1770-1771, 5'-1772-1773, 5'-1774-1775, 5'-1776-1777, 5'-1778-1779, 5'-1780-1781, 5'-1782-1783, 5'-1784-1785, 5'-1786-1787, 5'-1788-1789, 5'-1790-1791, 5'-1792-1793, 5'-1794-1795, 5'-1796-1797, 5'-1798-1799, 5'-1800-1801, 5'-1802-1803, 5'-1804-1805, 5'-1806-1807, 5'-1808-1809, 5'-1810-1811, 5'-1812-1813, 5'-1814-1815, 5'-1816-1817, 5'-1818-1819, 5'-1820-1821, 5'-1822-1823, 5'-1824-1825, 5'-1826-1827, 5'-1828-1829, 5'-1830-1831, 5'-1832-1833, 5'-1834-1835, 5'-1836-1837, 5'-1838-1839, 5'-1840-1841, 5'-1842-1843, 5'-1844-1845, 5'-1846-1847, 5'-1848-1849, 5'-1850-1851, 5'-1852-1853, 5'-1854-1855, 5'-1856-1857, 5'-1858-1859, 5'-1860-1861, 5'-1862-1863, 5'-1864-1865, 5'-1866-1867, 5'-1868-1869, 5'-1870-1871, 5'-1872-1873, 5'-1874-1875, 5'-1876-1877, 5'-1878-1879, 5'-1880-1881, 5'-1882-1883, 5'-1884-1885, 5'-1886-1887, 5'-1888-1889, 5'-1890-1891, 5'-1892-1893, 5'-1894-1895, 5'-1896-1897, 5'-1898-1899, 5'-1900-1901, 5'-1902-1903, 5'-1904-1905, 5'-1906-1907, 5'-1908-1909, 5'-1910-1911, 5'-1912-1913, 5'-1914-1915, 5'-1916-1917, 5'-1918-1919, 5'-1920-1921, 5'-1922-1923, 5'-1924-1925, 5'-1926-1927, 5'-1928-1929, 5'-1930-1931, 5'-1932-1933, 5'-1934-1935, 5'-1936-1937, 5'-1938-1939, 5'-1940-1941, 5'-1942-1943, 5'-1944-1945, 5'-1946-1947, 5'-1948-1949, 5'-1950-1951, 5'-1952-1953, 5'-1954-1955, 5'-1956-1957, 5'-1958-1959, 5'-1960-1961, 5'-1962-1963, 5'-1964-1965, 5'-1966-1967, 5'-1968-1969, 5'-1970-1971, 5'-1972-1973, 5'-1974-1975, 5'-1976-1977, 5'-1978-1979, 5'-1980-1981, 5'-1982-1983, 5'-1984-1985, 5'-1986-1987, 5'-1988-1989, 5'-1990-1991, 5'-1992-1993, 5'-1994-1995, 5'-1996-1997, 5'-1998-1999, 5'-2000-2001, 5'-2002-2003, 5'-2004-2005, 5'-2006-2007, 5'-2008-2009, 5'-2010-2011, 5'-2012-2013, 5'-2014-2015, 5'-2016-2017, 5'-2018-2019, 5'-202

A:Cross-references: EMBL:X84153
 R.Schadiner, C.; Leusch, H.; Hagmann, R.
 Submitted to the EMBL Data Library, January 1995
 A:Description: Detection of point mutations in chloroplast genes of *Antirrhinum majus* L.
 A:Reference number: 552394
 A:Accession: S52395
 A:Molecule type: DNA
 A:Residues: 1-566, 'T', 568-734, <SC2>
 A:Cross-references: EMBL:X84153, NID:q669095, PIDN:CAA58358.1, PID:q664096
 A:Accession: S52394
 A:Molecule type: DNA
 A:Residues: 1-135, <SC2>
 A:Cross-references: EMBL:X84152, NID:q660003, PIDN:CAA58357.1, PID:q660004
 A:Genetics:
 A:Gene: psbB
 A:Genome: chloroplast
 C:Suprafamily: photosystem 1 p700 apoprotein
 C:Keywords: chloroplast, electron transfer, membrane-associated complex; photosynthesis;
 Query Match 73.6%, Score 39; DB 2; Length 734;
 Best Local Similarity 75.0%; Pred. No. 25;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 GNFSAWV 9
 DB 71 GNFSAWV 78
 RESULT 11
 C:0821
 hypothetical protein ECS1539 [imported] - Escherichia coli (strain O157:H7, substrain R)
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C:Accession: C90821
 R.Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 Kasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen-
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: C90821
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-61, <HAY>
 A:Cross-references: CH:HA000007; PIDN:BAH4962.1; PID:q13361003; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RMD 050952
 C:Genetics:
 A:Gene: ECS1539
 Query Match 71.7%, Score 38; DB 2; Length 61;
 Best Local Similarity 85.7%; Pred. No. 3.2;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 4 GNFSAWV 9
 DB 45 GNFSAWV 41
 RESULT 12
 C:0842
 hypothetical protein ECS1628 [imported] - Escherichia coli (strain O157:H7, substrain R)
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C:Accession: D90832
 R.Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 Kasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen-
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: D90832
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-64, <HAY>
 A:Cross-references: GB:BA000007; PIDN:BAH35051.1; PID:q14461092; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RMD 050952
 C:Genetics:
 A:Gene: ECS1628

Query Match 71.7%, Score 38; DB 2; Length 64;
 Best Local Similarity 85.7%; Pred. No. 3.4;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 GNFSAWV 9
 DB 35 GNFSAWV 41

RESULT 13
 P85689
 unknown protein encoded by prophage CP_933X [imported] - Escherichia coli (strain O157:H7, substrain RMD933)
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C:Accession: P85689
 R.Ferna, N.T.; Plunkett III, G.; Burland, V.; Mau, H.; Glasner, J.D.; Rose, D.J.; May
 Miller, L.; Grobbeck, F.J.; Davis, N.W.; Lim, A.; Dolan, F.; Potamousis, K.; Apoda
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: P85689
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-64, <STO>
 A:Cross-references: GB:AE005174, NID:q12514809; PIDN:AA656978.1; GSPDB:GN00145; UNWIP;
 A:Experimental source: strain O157:H7, substrain RMD933
 C:Genetics:
 A:Gene: Z1881

Query Match 71.7%, Score 38; DB 2; Length 64;
 Best Local Similarity 85.7%; Pred. No. 3.4;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 GNFSAWV 9
 DB 35 GNFSAWV 41

RESULT 14
 E90901
 hypothetical protein ECS2181 [imported] - Escherichia coli (strain O157:H7, substrain RMD901)
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C:Accession: E90901
 R.Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
 Kasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and q
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: E90901
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-74, <HAY>
 A:Cross-references: GB:BA000007; PIDN:BAH35604.1; PID:q14461647; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RMD 050952
 C:Genetics:
 A:Gene: ECS2181

Query Match 71.7%, Score 38; DB 2; Length 74;
 Best Local Similarity 85.7%; Pred. No. 3.9;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 GNFSAWV 9
 DB 35 GNFSAWV 41

Db 35 GNFSAWV 41

RESULT 15

H85743
 unknown protein encoded within prophage CP 933R [imported] - Escherichia coli (strain O157:H7)
 C:Species: Escherichia coli
 C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: H85743
 R:Perin, N.T.; Plunkett III, G.; Burland, V.; Mau, R.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca, J.
 Nature 409, 529-533, 2001
 A:Title: Genome Sequence of Enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21974935; PMID:11206551
 A:Accession: H85743
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-74 <STO>
 A:Cross-references: GR:AP005174; NID:012515361; PIFN:MA056412 1; GNPB:GN00145; UWCP:223
 A:Experimental source: strain O157:H7, Substrain EDL933
 C:Genetics:
 A:Gene: 22366

Query Match 71.7%; Score 38; DB 2; Length 74;
 Best Local Similarity 85.7%; Pred No. 3.9;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 3 GNFSAWV 9
 |||:|
 Db 35 GNFSAWV 41

Search completed: September 5, 2002, 15:28:49
 Job time: 354 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein protein search, using sw model

Run on: Sept-04-02 5:20:02, 16:28:43 Search time 33.99 seconds
(without alignments)
10 252 million cell updates/sec

Title: US-09-744-804-35
Perfect score: 53
Sequence: 1 KQNFENAW 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 20000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	53	100.0	387	1	MFCM_HUMAN
2	42	79.2	734	1	PSAB_EUCCR
3	42	79.2	734	1	PSAB_GENHO
4	42	79.2	734	1	PSAB_PEA
5	42	79.2	734	1	PSAB_PINTH
6	41	77.4	734	1	PSAB_CYACA
7	39	73.6	369	1	CYRG_MOUSE
8	39	73.6	379	1	CYRG_BOVIN
9	39	73.6	734	1	PSAB_ANTMA
10	39	73.6	734	1	PSAB_SPTOL
11	39	73.6	734	1	PSAB_TSPAC
12	38	71.7	409	1	MFCM_PIG
13	38	71.7	427	1	MFCM_BOVIN
14	38	71.7	734	1	PSAB_ARATH
15	38	71.7	734	1	PSAB_MARPO
16	38	71.7	734	1	PSAB_ORYSA
17	38	71.7	734	1	PSAB_WHEAT
18	38	71.7	735	1	PSAB_MAIZE
19	38	71.7	994	1	BCAL_STAXY
20	37	69.8	734	1	PSAB_GPOSI
21	37	69.8	734	1	PSAB_CHLYV
22	37	69.8	734	1	PSAB_GHITH
23	37	69.8	734	1	PSAB_MESVI
24	37	69.8	735	1	PSAB_CHEMG
25	37	69.8	735	1	PSAB_CHEUR
26	37	69.8	737	1	PSAB_CYAPA
27	37	69.8	738	1	PSAB_SYNPH
28	37	69.8	740	1	PSAB_SYNPH
29	37	69.8	747	1	PSAB_IPOMA
30	36	67.9	477	1	CHVL_MOUSE
31	36	67.9	503	1	CHVL_MOUSE
32	36	67.9	508	1	CHVL_MOUSE
33	36	67.9	730	1	PSAB_SYNPH

ALIGNMENTS

RESULT 1

ID	MFCM_HUMAN	STANDARD	PRT	387 AA.
AC	Q08431			
DT	01-OCT-1996 (ref. 34, Created)			
DT	01-NOV-1997 (ref. 35, last sequence update)			
DT	16-OCT-2001 (ref. 40, last annotation update)			
DE	Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-EB) (HMEG)			
DE	(Breast epithelial antigen BA46) (MFCM) [Contains: Medin].			
GN	MFCB8.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID:9606			
ON	[1]			
PP	SEQUENCE FROM N.A.			
RC	TISSUE=Breast, and Breast carcinoma;			
RX	MEDLINE=96213908; PubMed=8639264;			
PA	Conte F P, Taylor M P, Gelin S G, Cerini P L, Peterson J A;			
PA	"Cloning and sequence analysis of human breast epithelial antigen			
PA	BA46 reveals an RGD cell adhesion sequence presented on an epidermal			
PA	growth factor-like domain."			
RL	DNA Cell Biol. 15:281-286(1996).			
RN	[2]			
RP	SEQUENCE OF 170 387 FROM N.A.			
RC	TISSUE=Mammary gland;			
EX	MEDLINE 91371251; PubMed 1999992;			
PA	Latouche P, Peterson J A, Gelin S G, Cerini P L,			
PA	Cerini P L;			
PA	"A Mr 46,000 human milk fat globule protein that is highly expressed			
PA	in human breast tumors contains factor VIII-like domains."			
PA	Cancer Res 51:4994-4998(1991).			
RN	[3]			
PP	FAPPIA: SEQUENCE, AND CHARACTERIZATION.			
RC	TISSUE=Milk;			
RX	MEDLINE=98194924; PubMed=9535276;			
PA	Giutrida M G, Cavalletto M, Giunta C, Conti A,			
PA	Gederae Zimmermann J;			
PA	"Isolation and characterization of full and truncated forms of human			
PA	breast carcinoma protein BA46 from human milk fat globule membranes."			
PA	J. Protein Chem. 17:143-148(1998).			
RN	[4]			
PP	SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.			
RX	MEDLINE=99443076; PubMed=10411933;			
PA	Bachmann B, Raschke J, Schellen F, Westermarck G, Macchiano G,			
PA	Tjernberg L O, Nordstedt C, Engstrom U, Westermarck P;			
PA	"Medin, an internal fragment of acidic sheddase, muscle cell-produced			
PA	Lactadherin forms the most common human amyloid."			
PA	Proc Natl Acad Sci USA 96:8689-8694(1999).			
RN	[5]			
PP	CHARACTERIZATION.			
RX	MEDLINE 9745595; PubMed 9269429;			
PA	Taylor M P, Conte F P, Cerini P L, Peterson J A;			
PA	"Lactadherin (formerly BA46), a membrane associated glycoprotein			
PA	expressed in human milk and breast carcinomas, promotes A9.3; Asp			
PA	(95k)-dependent cell adhesion."			

RL DNA Cell Biol. 16:861-869(1997).
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
 CC SPECIFICALLY TO ROTAVIRUS AND INHIBITS ITS REPLICATION.
 CC -1- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF ACROTIC MEDIAN. AMYLOID
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND ACROTIC
 CC MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
 CC -1- PTM: MEDIN HAS A PACHEP N-TERMINUS WITH MINOP SPECIFS STARTING AT
 CC AMINO ACID 264 AND 273.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE D'DMAIN
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS
 CC This SWISS-prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch)

EMBL: U58516; AAC50549.1; -
 EMBL: S56151; AAB19771.1; -
 MIM: 602281; -
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000421; FA5R_C
 DR Pfam: PF00608; EGF_1
 DR Pfam: PF00754; F5_P8_type_C; 2.
 DR SMART: SM00181; EGF_1.
 DR SMART: SM00241; FA5R_C; 2.
 DR PROSITE: PS00222; EGF_1; 1.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01285; FA5R_C; 2.
 DR PROSITE: PS01286; FA5R_C; 2.
 DR Signal: Glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
 KW SIGNAL.
 FT CHAIN 24 487 LACTADHERIN.
 FT CHAIN 202 387 LACTADHERIN: SHORT FORM
 FT CHAIN 268 317 MEDIN.
 FT DOMAIN 24 67 EGF LIKE.
 FT DOMAIN 70 225 F5/8 TYPE C 1.
 FT DOMAIN 240 387 F5/8 TYPE C 2.
 FT SITE 46 48 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 27 38 BY SIMILARITY.
 FT DISULFID 32 55 BY SIMILARITY.
 FT DISULFID 57 66 BY SIMILARITY.
 FT DISULFID 70 225 BY SIMILARITY.
 FT DISULFID 212 216 BY SIMILARITY.
 FT DISULFID 230 387 BY SIMILARITY.
 FT CARBOHYD 238 238 N LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 325 325 N LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 329 329 N LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 350 350 N LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 487 AA; 4125 MW; 2EF55710F5A578210 CPD64;

Query Match 100.0%; Score 53; DR 1; Length 387;
 Best Local Similarity 100.0%; Pred. No. 0.01;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QCNFNAMW 9
 |||||

DB 271 QCNFNAMW 279

RESULT 2

ID PSAB_LOLJA STANDARD; PRT; 734 AA.

AC P58485;

DT 01-MAR-2002 (rel. 41, Created)

DI 01-MAR-2002 (rel. 41, Last sequence update)

DE Photosystem I P700 chlorophyll A apoprotein A2 (psab) (PST-B).

GN PSAB.

OS Lotus japonicus.
 OG Chloroplast.
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 1;
 CC Fabales; Fabaceae; Papilionoideae; Lotus.
 CC NCBI_TaxID=34305;
 RN |||
 PP SPOUNCE FROM N.A.
 RC STRAIN Accession MC-20;
 EX MEDLINE=21082929; PubMed=11214967;
 RA Kato T., Kaneko T., Sato S., Nakamura Y., Tabata S.;
 RT "Complete structure of the chloroplast genome of a legume, Lotus
 japonicus."
 RL DNA Res. 7:323-330(2000).
 CC -1- FUNCTION: psaa and psab bind P700, the primary electron donor of
 CC photosystem I (PSI), as well as the electron acceptors A0, A1, and
 CC FX. PSI functions as a plastocyanin-binding site.
 CC oxidoreductase
 CC -1- COFACTOR: P700 is a chlorophyll A dimer, A0 is chlorophyll A, A1
 CC is a phytylquinone and FX is a 4Fe-4S iron sulfur center.
 CC -1- SUBUNIT: A psaa/B heterodimer binds the P700 chlorophyll special
 CC pair and subsequent electron acceptors. The PSI reaction center of
 CC higher plants and algae is composed of one at least 11 subunits.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
 CC thylakoid membrane.
 CC -1- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.

This SWISS-prot entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch)

EMBL: AF002983; BAB33187.1; -
 PROSITE: PS00419; PHOTOSYSTEM_I; PSAB; 1.
 KW Photosynthesis; Photosystem I; Electron transport; Chloroplast;
 KW Thylakoid; Transmembrane, Iron-sulfur, 4Fe-4S, Chlorophyll.
 FT TRANSMEM 46 69 I (POTENTIAL).
 FT TRANSMEM 135 158 II (POTENTIAL).
 FT TRANSMEM 175 199 III (POTENTIAL).
 FT TRANSMEM 273 291 IV (POTENTIAL).
 FT TRANSMEM 330 353 V (POTENTIAL).
 FT TRANSMEM 369 395 VI (POTENTIAL).
 FT TRANSMEM 417 439 VII (POTENTIAL).
 FT TRANSMEM 517 535 VIII (POTENTIAL).
 FT TRANSMEM 575 596 IX (POTENTIAL).
 FT TRANSMEM 643 665 X (POTENTIAL).
 FT TRANSMEM 707 727 XI (POTENTIAL).
 FT METAL 559 559 IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
 FT METAL 568 568 IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
 FT BINDING 654 654 P700 SPECIAL PAIR CHLOROPHYLL AXIAL
 FT BINDING 662 662 LIGAND (BY SIMILARITY).
 FT BINDING 670 670 A0 CHLOROPHYLL (BY SIMILARITY).
 FT BINDING 671 671 A1 CHLOROPHYLL (BY SIMILARITY).
 FT BINDING 700 700 A1 PHYTYLQUINONE (BY SIMILARITY).
 SQ SEQUENCE 734 AA; 82362 MW; F83F2357E94BDF6 CPD64;

Query Match 79.2%; Score 42; DR 1; Length 744;
 Best Local Similarity 87.5%; Pred. No. 2.5;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 QCNFNAMW 9

DB 71 QCNFNAMW 78

RESULT 3

PSAB_OENHO

ID PSAB_OENHO STANDARD; PRT; 734 AA.

AC O9MTN7;
 DT 01-MAR-2002 (rel. 41, created)
 DT 01-MAR-2002 (rel. 41, last sequence update)
 DT 01-MAR-2002 (rel. 41, last annotation update)
 DE Photosystem I P700 chlorophyll A apoprotein A2 (psaA) (PST-I-B).
 GN PSAB.
 OS Oenothera lutea (Hooker's evening primrose).
 OG Chloroplast.
 CC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 CC Malvaceae; Onagraceae; Euphorbiaceae; Rosaceae; Euphorbiaceae; Rosaceae;
 CC Myricales; Gnaphalaceae; Oenothera.
 OX NCBI_TaxID=85636;
 RN [1]
 RP SEQUENCE FROM N A
 RC SIFALIN-CV Johanson.
 RX MEDLINE=2006978; PubMed=10952479;
 RA Hupfer H., Swiatek M., Hornung S., Herrmann R.G., Maier R.M.,
 RA Chiu W.-L., Sears R.
 RT "Complete nucleotide sequence of the Oenothera elata plastid
 RT chloroplast genome, representing plastome I of the five distinguishable
 RI Oenothera plastomes".
 RL Mol. Gen. Genet. 263:561-585 (2000)
 CC -!- FUNCTION: PsaA and psaB bind P700, the primary electron donor of
 CC photosystem I (PSI), as well as the electron acceptors A0, A1, and
 CC FX. PSI functions as a plastocyanin/cytochrome c6 ferredoxin
 CC oxidoreductase.
 CC -!- COFACTOR: P700 is a chlorophyll A dimer, A0 is chlorophyll A, A1
 CC is a phytylquinone and FX is a 4Fe-4S iron-sulfur center.
 CC -!- SUBUNIT: A psaA/B heterodimer binds the P700 chlorophyll special
 CC pair and subsequent electron acceptors. The PSI reaction center of
 CC higher plants and algae is composed of one at least 11 subunits.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
 CC thylakoid membrane.
 CC -!- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (<http://www.isb.ch>)
 CC or send an email to license@isb-sib.ch.
 CC
 CC EMBL: A2271079; CAB67138.1;
 CC InterPro: IPR001280; Psaa_psaB.
 CC Pfam: PF00223; psaa_psaB; 1.
 CC PRINTS: PR00257; PROTSYSPSAAB.
 CC PROSITE: PS00419; PHOTOSYSTEM_I_PSAAB; 1
 CC Phylloquinone; Photosystem I; Ferredoxin; Transmembrane; Chloroplast;
 CC Thylakoid; Transmembrane; Iron-sulfur; 4Fe-4S; Chlorophyll.
 CC TRANSMEM 46 69
 CC TRANSMEM 135 157
 CC TRANSMEM 175 199
 CC TRANSMEM 273 291
 CC TRANSMEM 330 353
 CC TRANSMEM 369 395
 CC TRANSMEM 417 439
 CC TRANSMEM 517 535
 CC TRANSMEM 575 596
 CC TRANSMEM 643 665
 CC TRANSMEM 707 727
 CC TRANSMEM 767 789
 CC TRANSMEM 843 865
 CC TRANSMEM 927 949
 CC TRANSMEM 1011 1033
 CC TRANSMEM 1095 1117
 CC TRANSMEM 1199 1221
 CC TRANSMEM 1303 1325
 CC TRANSMEM 1407 1429
 CC TRANSMEM 1511 1533
 CC TRANSMEM 1615 1637
 CC TRANSMEM 1719 1741
 CC TRANSMEM 1823 1845
 CC TRANSMEM 1927 1949
 CC TRANSMEM 2031 2053
 CC TRANSMEM 2135 2157
 CC TRANSMEM 2239 2261
 CC TRANSMEM 2343 2365
 CC TRANSMEM 2447 2469
 CC TRANSMEM 2551 2573
 CC TRANSMEM 2655 2677
 CC TRANSMEM 2759 2781
 CC TRANSMEM 2863 2885
 CC TRANSMEM 2967 2989
 CC TRANSMEM 3071 3093
 CC TRANSMEM 3175 3197
 CC TRANSMEM 3279 3301
 CC TRANSMEM 3383 3405
 CC TRANSMEM 3487 3509
 CC TRANSMEM 3611 3633
 CC TRANSMEM 3715 3737
 CC TRANSMEM 3819 3841
 CC TRANSMEM 3923 3945
 CC TRANSMEM 4027 4049
 CC TRANSMEM 4131 4153
 CC TRANSMEM 4235 4257
 CC TRANSMEM 4339 4361
 CC TRANSMEM 4443 4465
 CC TRANSMEM 4547 4569
 CC TRANSMEM 4651 4673
 CC TRANSMEM 4755 4777
 CC TRANSMEM 4859 4881
 CC TRANSMEM 4963 4985
 CC TRANSMEM 5067 5089
 CC TRANSMEM 5171 5193
 CC TRANSMEM 5275 5297
 CC TRANSMEM 5379 5401
 CC TRANSMEM 5483 5505
 CC TRANSMEM 5587 5609
 CC TRANSMEM 5711 5733
 CC TRANSMEM 5815 5837
 CC TRANSMEM 5919 5941
 CC TRANSMEM 6023 6045
 CC TRANSMEM 6127 6149
 CC TRANSMEM 6231 6253
 CC TRANSMEM 6335 6357
 CC TRANSMEM 6439 6461
 CC TRANSMEM 6543 6565
 CC TRANSMEM 6647 6669
 CC TRANSMEM 6751 6773
 CC TRANSMEM 6855 6877
 CC TRANSMEM 6959 6981
 CC TRANSMEM 7063 7085
 CC TRANSMEM 7167 7189
 CC TRANSMEM 7271 7293
 CC TRANSMEM 7375 7397
 CC TRANSMEM 7479 7501
 CC TRANSMEM 7583 7605
 CC TRANSMEM 7687 7709
 CC TRANSMEM 7791 7813
 CC TRANSMEM 7895 7917
 CC TRANSMEM 7999 8021
 CC TRANSMEM 8103 8125
 CC TRANSMEM 8207 8229
 CC TRANSMEM 8311 8333
 CC TRANSMEM 8415 8437
 CC TRANSMEM 8519 8541
 CC TRANSMEM 8623 8645
 CC TRANSMEM 8727 8749
 CC TRANSMEM 8831 8853
 CC TRANSMEM 8935 8957
 CC TRANSMEM 9039 9061
 CC TRANSMEM 9143 9165
 CC TRANSMEM 9247 9269
 CC TRANSMEM 9351 9373
 CC TRANSMEM 9455 9477
 CC TRANSMEM 9559 9581
 CC TRANSMEM 9663 9685
 CC TRANSMEM 9767 9789
 CC TRANSMEM 9871 9893
 CC TRANSMEM 9975 9997
 CC TRANSMEM 10079 10101
 CC TRANSMEM 10183 10205
 CC TRANSMEM 10287 10309
 CC TRANSMEM 10391 10413
 CC TRANSMEM 10495 10517
 CC TRANSMEM 10599 10621
 CC TRANSMEM 10693 10715
 CC TRANSMEM 10797 10819
 CC TRANSMEM 10901 10923
 CC TRANSMEM 11005 11027
 CC TRANSMEM 11109 11131
 CC TRANSMEM 11213 11235
 CC TRANSMEM 11317 11339
 CC TRANSMEM 11421 11443
 CC TRANSMEM 11525 11547
 CC TRANSMEM 11629 11651
 CC TRANSMEM 11733 11755
 CC TRANSMEM 11837 11859
 CC TRANSMEM 11941 11963
 CC TRANSMEM 12045 12067
 CC TRANSMEM 12149 12171
 CC TRANSMEM 12253 12275
 CC TRANSMEM 12357 12379
 CC TRANSMEM 12461 12483
 CC TRANSMEM 12565 12587
 CC TRANSMEM 12669 12691
 CC TRANSMEM 12773 12795
 CC TRANSMEM 12877 12899
 CC TRANSMEM 12981 13003
 CC TRANSMEM 13085 13107
 CC TRANSMEM 13189 13211
 CC TRANSMEM 13293 13315
 CC TRANSMEM 13397 13419
 CC TRANSMEM 13501 13523
 CC TRANSMEM 13605 13627
 CC TRANSMEM 13709 13731
 CC TRANSMEM 13813 13835
 CC TRANSMEM 13917 13939
 CC TRANSMEM 14021 14043
 CC TRANSMEM 14125 14147
 CC TRANSMEM 14229 14251
 CC TRANSMEM 14333 14355
 CC TRANSMEM 14437 14459
 CC TRANSMEM 14541 14563
 CC TRANSMEM 14645 14667
 CC TRANSMEM 14749 14771
 CC TRANSMEM 14853 14875
 CC TRANSMEM 14957 14979
 CC TRANSMEM 15061 15083
 CC TRANSMEM 15165 15187
 CC TRANSMEM 15269 15291
 CC TRANSMEM 15373 15395
 CC TRANSMEM 15477 15499
 CC TRANSMEM 15581 15603
 CC TRANSMEM 15685 15707
 CC TRANSMEM 15789 15811
 CC TRANSMEM 15893 15915
 CC TRANSMEM 15997 16019
 CC TRANSMEM 16101 16123
 CC TRANSMEM 16205 16227
 CC TRANSMEM 16309 16331
 CC TRANSMEM 16413 16435
 CC TRANSMEM 16517 16539
 CC TRANSMEM 16621 16643
 CC TRANSMEM 16725 16747
 CC TRANSMEM 16829 16851
 CC TRANSMEM 16933 16955
 CC TRANSMEM 17037 17059
 CC TRANSMEM 17141 17163
 CC TRANSMEM 17245 17267
 CC TRANSMEM 17349 17371
 CC TRANSMEM 17453 17475
 CC TRANSMEM 17557 17579
 CC TRANSMEM 17661 17683
 CC TRANSMEM 17765 17787
 CC TRANSMEM 17869 17891
 CC TRANSMEM 17973 17995
 CC TRANSMEM 18077 18099
 CC TRANSMEM 18181 18203
 CC TRANSMEM 18285 18307
 CC TRANSMEM 18389 18411
 CC TRANSMEM 18493 18515
 CC TRANSMEM 18597 18619
 CC TRANSMEM 18691 18713
 CC TRANSMEM 18795 18817
 CC TRANSMEM 18899 18921
 CC TRANSMEM 18993 19015
 CC TRANSMEM 19097 19119
 CC TRANSMEM 19191 19213
 CC TRANSMEM 19295 19317
 CC TRANSMEM 19399 19421
 CC TRANSMEM 19493 19515
 CC TRANSMEM 19597 19619
 CC TRANSMEM 19691 19713
 CC TRANSMEM 19795 19817
 CC TRANSMEM 19899 19921
 CC TRANSMEM 19993 20015
 CC TRANSMEM 20097 20119
 CC TRANSMEM 20191 20213
 CC TRANSMEM 20295 20317
 CC TRANSMEM 20399 20421
 CC TRANSMEM 20493 20515
 CC TRANSMEM 20597 20619
 CC TRANSMEM 20691 20713
 CC TRANSMEM 20795 20817
 CC TRANSMEM 20899 20921
 CC TRANSMEM 20993 21015
 CC TRANSMEM 21097 21119
 CC TRANSMEM 21191 21213
 CC TRANSMEM 21295 21317
 CC TRANSMEM 21399 21421
 CC TRANSMEM 21493 21515
 CC TRANSMEM 21597 21619
 CC TRANSMEM 21691 21713
 CC TRANSMEM 21795 21817
 CC TRANSMEM 21899 21921
 CC TRANSMEM 21993 22015
 CC TRANSMEM 22097 22119
 CC TRANSMEM 22191 22213
 CC TRANSMEM 22295 22317
 CC TRANSMEM 22399 22421
 CC TRANSMEM 22493 22515
 CC TRANSMEM 22597 22619
 CC TRANSMEM 22691 22713
 CC TRANSMEM 22795 22817
 CC TRANSMEM 22899 22921
 CC TRANSMEM 22993 23015
 CC TRANSMEM 23097 23119
 CC TRANSMEM 23191 23213
 CC TRANSMEM 23295 23317
 CC TRANSMEM 23399 23421
 CC TRANSMEM 23493 23515
 CC TRANSMEM 23597 23619
 CC TRANSMEM 23691 23713
 CC TRANSMEM 23795 23817
 CC TRANSMEM 23899 23921
 CC TRANSMEM 23993 24015
 CC TRANSMEM 24097 24119
 CC TRANSMEM 24191 24213
 CC TRANSMEM 24295 24317
 CC TRANSMEM 24399 24421
 CC TRANSMEM 24493 24515
 CC TRANSMEM 24597 24619
 CC TRANSMEM 24691 24713
 CC TRANSMEM 24795 24817
 CC TRANSMEM 24899 24921
 CC TRANSMEM 24993 25015
 CC TRANSMEM 25097 25119
 CC TRANSMEM 25191 25213
 CC TRANSMEM 25295 25317
 CC TRANSMEM 25399 25421
 CC TRANSMEM 25493 25515
 CC TRANSMEM 25597 25619
 CC TRANSMEM 25691 25713
 CC TRANSMEM 25795 25817
 CC TRANSMEM 25899 25921
 CC TRANSMEM 25993 26015
 CC TRANSMEM 26097 26119
 CC TRANSMEM 26191 26213
 CC TRANSMEM 26295 26317
 CC TRANSMEM 26399 26421
 CC TRANSMEM 26493 26515
 CC TRANSMEM 26597 26619
 CC TRANSMEM 26691 26713
 CC TRANSMEM 26795 26817
 CC TRANSMEM 26899 26921
 CC TRANSMEM 26993 27015
 CC TRANSMEM 27097 27119
 CC TRANSMEM 27191 27213
 CC TRANSMEM 27295 27317
 CC TRANSMEM 27399 27421
 CC TRANSMEM 27493 27515
 CC TRANSMEM 27597 27619
 CC TRANSMEM 27691 27713
 CC TRANSMEM 27795 27817
 CC TRANSMEM 27899 27921
 CC TRANSMEM 27993 28015
 CC TRANSMEM 28097 28119
 CC TRANSMEM 28191 28213
 CC TRANSMEM 28295 28317
 CC TRANSMEM 28399 28421
 CC TRANSMEM 28493 28515
 CC TRANSMEM 28597 28619
 CC TRANSMEM 28691 28713
 CC TRANSMEM 28795 28817
 CC TRANSMEM 28899 28921
 CC TRANSMEM 28993 29015
 CC TRANSMEM 29097 29119
 CC TRANSMEM 29191 29213
 CC TRANSMEM 29295 29317
 CC TRANSMEM 29399 29421
 CC TRANSMEM 29493 29515
 CC TRANSMEM 29597 29619
 CC TRANSMEM 29691 29713
 CC TRANSMEM 29795 29817
 CC TRANSMEM 29899 29921
 CC TRANSMEM 29993 30015
 CC TRANSMEM 30097 30119
 CC TRANSMEM 30191 30213
 CC TRANSMEM 30295 30317
 CC TRANSMEM 30399 30421
 CC TRANSMEM 30493 30515
 CC TRANSMEM 30597 30619
 CC TRANSMEM 30691 30713
 CC TRANSMEM 30795 30817
 CC TRANSMEM 30899 30921
 CC TRANSMEM 30993 31015
 CC TRANSMEM 31097 31119
 CC TRANSMEM 31191 31213
 CC TRANSMEM 31295 31317
 CC TRANSMEM 31399 31421
 CC TRANSMEM 31493 31515
 CC TRANSMEM 31597 31619
 CC TRANSMEM 31691 31713
 CC TRANSMEM 31795 31817
 CC TRANSMEM 31899 31921
 CC TRANSMEM 31993 32015
 CC TRANSMEM 32097 32119
 CC TRANSMEM 32191 32213
 CC TRANSMEM 32295 32317
 CC TRANSMEM 32399 32421
 CC TRANSMEM 32493 32515
 CC TRANSMEM 32597 32619
 CC TRANSMEM 32691 32713
 CC TRANSMEM 32795 32817
 CC TRANSMEM 32899 32921
 CC TRANSMEM 32993 33015
 CC TRANSMEM 33097 33119
 CC TRANSMEM 33191 33213
 CC TRANSMEM 33295 33317
 CC TRANSMEM 33399 33421
 CC TRANSMEM 33493 33515
 CC TRANSMEM 33597 33619
 CC TRANSMEM 33691 33713
 CC TRANSMEM 33795 33817
 CC TRANSMEM 33899 33921
 CC TRANSMEM 33993 34015
 CC TRANSMEM 34097 34119
 CC TRANSMEM 34191 34213
 CC TRANSMEM 34295 34317
 CC TRANSMEM 34399 34421
 CC TRANSMEM 34493 34515
 CC TRANSMEM 34597 34619
 CC TRANSMEM 34691 34713
 CC TRANSMEM 34795 34817
 CC TRANSMEM 34899 34921
 CC TRANSMEM 34993 35015
 CC TRANSMEM 35097 35119
 CC TRANSMEM 35191 35213
 CC TRANSMEM 35295 35317
 CC TRANSMEM 35399 35421
 CC TRANSMEM 35493 35515
 CC TRANSMEM 35597 35619
 CC TRANSMEM 35691 35713
 CC TRANSMEM 35795 35817
 CC TRANSMEM 35899 35921
 CC TRANSMEM 35993 36015
 CC TRANSMEM 36097 36119
 CC TRANSMEM 36191 36213
 CC TRANSMEM 36295 36317
 CC TRANSMEM 36399 36421
 CC TRANSMEM 36493 36515
 CC TRANSMEM 36597 36619
 CC TRANSMEM 36691 36713
 CC TRANSMEM 36795 36817
 CC TRANSMEM 36899 36921
 CC TRANSMEM 36993 37015
 CC TRANSMEM 37097 37119
 CC TRANSMEM 37191 37213
 CC TRANSMEM 37295 37317
 CC TRANSMEM 37399 37421
 CC TRANSMEM 37493 37515
 CC TRANSMEM 37597 37619
 CC TRANSMEM 37691 37713
 CC TRANSMEM 37795 37817
 CC TRANSMEM 37899 37921
 CC TRANSMEM 37993 38015
 CC TRANSMEM 38097 38119
 CC TRANSMEM 38191 38213
 CC TRANSMEM 38295 38317
 CC TRANSMEM 38399 38421
 CC TRANSMEM 38493 38515
 CC TRANSMEM 38597 38619
 CC TRANSMEM 38691 38713
 CC TRANSMEM 38795 38817
 CC TRANSMEM 38899 38921
 CC TRANSMEM 38993 39015
 CC TRANSMEM 39097 39119
 CC TRANSMEM 39191 39213
 CC TRANSMEM 39295 39317
 CC TRANSMEM 39399 39421
 CC TRANSMEM 39493 39515
 CC TRANSMEM 39597 39619
 CC TRANSMEM 39691 39713
 CC TRANSMEM 39795 39817
 CC TRANSMEM 39899 39921
 CC TRANSMEM 39993 40015
 CC TRANSMEM 40097 40119
 CC TRANSMEM 40191 40213
 CC TRANSMEM 40295 40317
 CC TRANSMEM 40399 40421
 CC TRANSMEM 40493 40515
 CC TRANSMEM 40597 40619
 CC TRANSMEM 40691 40713
 CC TRANSMEM 40795 40817
 CC TRANSMEM 40899 40921
 CC TRANSMEM 40993 41015
 CC TRANSMEM 41097 41119
 CC TRANSMEM 41191 41213
 CC TRANSMEM 41295 41317
 CC TRANSMEM 41399 41421
 CC TRANSMEM 41493 41515
 CC TRANSMEM 41597 41619
 CC TRANSMEM 41691 41713
 CC TRANSMEM 41795 41817
 CC TRANSMEM 41899 41921
 CC TRANSMEM 41993 42015
 CC TRANSMEM 42097 42119
 CC TRANSMEM 42191 42213
 CC TRANSMEM 42295 42317
 CC TRANSMEM 42399 42421
 CC TRANSMEM 42493 42515
 CC TRANSMEM 42597 42619
 CC TRANSMEM 42691 42713
 CC TRANSMEM 42795 42817
 CC TRANSMEM 42899 42921
 CC TRANSMEM 42993 43015
 CC TRANSMEM 43097 43119
 CC TRANSMEM 43191 43213
 CC TRANSMEM 43295 43317
 CC TRANSMEM 43399 43421
 CC TRANSMEM 43493 43515
 CC TRANSMEM 43597 43619
 CC TRANSMEM 43691 43713
 CC TRANSMEM 43795 43817
 CC TRANSMEM 43899 43921
 CC TRANSMEM 43993 44015
 CC TRANSMEM 44097 44119
 CC TRANSMEM 44191 44213
 CC TRANSMEM 44295 44317
 CC TRANSMEM 44399 44421
 CC TRANSMEM 44493 44515
 CC TRANSMEM 44597 44619
 CC TRANSMEM 44691 44713
 CC TRANSMEM 44795 44817
 CC TRANSMEM 44899 44921
 CC TRANSMEM 44993 45015
 CC TRANSMEM 45097 45119
 CC TRANSMEM 45191 45213
 CC TRANSMEM 45295 45317
 CC TRANSMEM 45399 45421
 CC TRANSMEM 45493 45515
 CC TRANSMEM 45597 45619
 CC TRANSMEM 45691 45713
 CC TRANSMEM 45795 45817
 CC TRANSMEM 45899 45921
 CC TRANSMEM 45993 46015
 CC TRANSMEM 46097 46119
 CC TRANSMEM 46191 46213
 CC TRANSMEM 46295 46317
 CC TRANSMEM 46399 46421
 CC TRANSMEM 46493 46515
 CC TRANSMEM 46597 46619
 CC TRANSMEM 46691 46713
 CC TRANSMEM 46795 46817
 CC TRANSMEM 46899 46921
 CC TRANSMEM 46993 47015
 CC TRANSMEM 47097 47119
 CC TRANSMEM 47191 47213
 CC TRANSMEM 47295 47317
 CC TRANSMEM 47399 47421
 CC TRANSMEM 47493 47515
 CC TRANSMEM 47597 47619
 CC TRANSMEM 47691 47713
 CC TRANSMEM 47795 47817
 CC TRANSMEM 47899 47921
 CC TRANSMEM 47993 48015
 CC TRANSMEM 48097 48119
 CC TRANSMEM 48191 48213
 CC TRANSMEM 48295 48317
 CC TRANSMEM 48399 48421
 CC TRANSMEM 48493 48515
 CC TRANSMEM 48597 48619
 CC TRANSMEM 48691 48713
 CC TRANSMEM 48795 48817
 CC TRANSMEM 48899 48921
 CC TRANSMEM 48993 49015
 CC TRANSMEM 49097 49119
 CC TRANSMEM 49191 49213
 CC TRANSMEM 49295 49317
 CC TRANSMEM 49399 49421
 CC TRANSMEM 49493 49515
 CC TRANSMEM 49597 49619
 CC TRANSMEM 49691 49713
 CC TRANSMEM 49795 49817
 CC TRANSMEM 49899 49921
 CC TRANSMEM 49993 50015
 CC TRANSMEM 50097 50119
 CC TRANSMEM 50191 50213
 CC TRANSMEM 50295 50317
 CC TRANSMEM 50399 50421
 CC TRANSMEM 50493 50515
 CC TRANSMEM 50597 50619
 CC TRANSMEM 50691 50713
 CC TRANSMEM 50795 50817
 CC TRANSMEM 50899 50921
 CC TRANSMEM 50993 51015
 CC TRANSMEM 51097 51119
 CC TRANSMEM 51191 51213
 CC TRANSMEM 51295 51317
 CC TRANSMEM 51399 51421
 CC TRANSMEM 51493 51515
 CC TRANSMEM 51597 51619
 CC TRANSMEM 51691 51713
 CC TRANSMEM 51795 51817
 CC TRANSMEM 51899 51921
 CC TRANSMEM 51993 52015
 CC TRANSMEM 52097 52119
 CC TRANSMEM 52191 52213
 CC TRANSMEM 52295 52317
 CC TRANSMEM 52399 52421
 CC TRANSMEM 52493 52515
 CC TRANSMEM 52597 52619
 CC TRANSMEM 52691 52713
 CC TRANSMEM 52795 52817
 CC TRANSMEM 52899 52921
 CC TRANSMEM 52993 53015
 CC TRANSMEM 53097 53119
 CC TRANSMEM 53191 53213
 CC TRANSMEM 53295 53317
 CC TRANSMEM 53399 53421
 CC TRANSMEM 53493 53515
 CC TRANSMEM 53597 53619
 CC TRANSMEM 53691 53713
 CC TRANSMEM 53795 53817
 CC TRANSMEM 53899 53921
 CC TRANSMEM 53993 54015
 CC TRANSMEM 54097 54119
 CC TRANSMEM 54191 54213
 CC TRANSMEM 54295 54317
 CC TRANSMEM 54399 54421
 CC TRANSMEM 54493 54515
 CC TRANSMEM 54597 54619
 CC TRANSMEM 54691 54713
 CC TRANSMEM 54795 54817
 CC TRANSMEM 54899 54921
 CC TRANSMEM 54993 55015
 CC TRANSMEM 55097 55119
 CC TRANSMEM 55191 55213
 CC TRANSMEM 55295 55317
 CC TRANSMEM 55399 55421
 CC TRANSMEM 55493 55515
 CC TRANSMEM 55597 55619
 CC TRANSMEM 55691 55713
 CC TRANSMEM 55795 55817
 CC TRANSMEM 55899 55921
 CC TRANSMEM 55993 56015
 CC TRANSMEM 56097 56119
 CC TRANSMEM 56191 56213
 CC TRANSMEM 56295 56317
 CC TRANSMEM 56399 56421
 CC TRANSMEM 56493 56515
 CC TRANSMEM 56597 56619
 CC TRANSMEM 56691 56713
 CC TRANSMEM 56795 56817
 CC TRANSMEM 56899 56921
 CC TRANSMEM 56993 57015
 CC TRANSMEM 57097 57119
 CC TRANSMEM 57191 57213
 CC TRANSMEM 57295 57317
 CC TRANSMEM 57399 57421
 CC TRANSMEM 57493 57515
 CC TRANSMEM 57597 57619
 CC TRANSMEM 57691 57713
 CC TRANSMEM 57795 57817
 CC TRANSMEM 57899 57921
 CC TRANSMEM 57993 58015
 CC TRANSMEM 58097 58119
 CC TRANSMEM 58191 58213
 CC TRANSMEM 58295 58317
 CC TRANSMEM 58399 58421
 CC TRANSMEM 58493 58515
 CC TRANSMEM 58597 58619
 CC TRANSMEM 58691 58713
 CC TRANSMEM 58795 58817
 CC TRANSMEM 58899 58921
 CC TRANSMEM 58993 59015
 CC TRANSMEM 59097 59119
 CC TRANSMEM 59191 59213
 CC TRANSMEM 59295 59317
 CC TRANSMEM 59399 59421
 CC TRANSMEM 59493 59515
 CC TRANSMEM 59597 59619
 CC TRANSMEM 59691 59713
 CC TRANSMEM 59795 59817
 CC TRANSMEM 59899 59921
 CC TRANSMEM 59993 60015
 CC TRANSMEM 60097 60119
 CC TRANSMEM 60191 6021

EMBL; AF022196; AAF12881.1; -
InterPro; IP0001280; PsaA_psaA
Pfam; PF00223; PsaA_psaB

Query Match 73.6%; Score 39; DB 1; Length 369;
 Best Local Similarity 85.7%; Pred. No. 4.5;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QGNFNAW 8
 |||||
 DB 304 QGNFSAW 310

RESULT 8
 ID CYRG_BOVIN STANDARD; PRT; 379 AA.
 AC Q95118;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DI 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Cytokine receptor common gamma chain precursor (Gamma-C)
 DE (Interleukin-2 receptor gamma chain) (IL 2R gamma chain) (P64).
 GN IL2RC.
 OS Bos taurus (Bovine)
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9626473; PubMed=8672241;
 RA Yoo J., Stone R.T., Solinas-Toldo S., Fries R., Beattie C.W.;
 FT "Cloning and chromosomal mapping of bovine interleukin-2 receptor
 gamma gene.";
 RL DNA Cell Biol. 15:453-459(1996).
 CC -!- FUNCTION: COMMON SUBUNIT FOR THE RECEPTORS FOR A VARIETY OF
 INTERLEUKINS
 CC -!- SUBUNIT: THE GAMMA CHAIN IS COMMON TO THE IL-2, IL-4, IL-7 AND
 CC PROBABLY ALSO THE IL-13 RECEPTORS.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.
 CC -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U33748; AA07812.1; -;
 DR HSSP: P31785; LILN.
 DR InterPro: IPR002496; CRFA.
 DR InterPro: IPR003461; FN_IIL.
 DR InterPro: IPR003531; Hematopo_receptor_S_F1.
 DR SMART: SM00041; I01; 1.
 DR SMART: SM00060; FN3; 1.
 DR PROSITE: PS01455; HPMAT040_PPC_S_F1; 1.
 KW Receptor; Transmembrane; Glycoprotein; Signal.
 FT SIGNAL: 1 22 POTENTIAL.
 FT CHAIN: 23 379
 FT DOMAIN: 23 269 EXTRACELLULAR (POTENTIAL).
 FT TRANSFM: 270 390 POTENTIAL.
 FT DOMAIN: 291 379
 FT DOMAIN: 158 256 CYTOPLASMIC (POTENTIAL).
 FT DISULFID: 68 78 FIBRONECTIN TYPE-III.
 FT DISULFID: 109 122 POTENTIAL.
 FT CARBOHYD: 77 77 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD: 81 81 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD: 90 90 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD: 166 166 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD: 171 171 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 379 AA. 430-47 MW. 33CFAD9C9BQ32178 CRC64;

Query Match 73.6%; Score 39; DB 1; Length 379;
 Best Local Similarity 85.7%; Pred. No. 4.6;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QGNFNAW 8
 |||||
 DB 317 QGNFSAW 323

RESULT 9
 ID PSAB_ANTMA STANDARD; PRT; 734 AA.
 AC Q33432;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DI 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Photosystem I p700 chlorophyll A apoprotein A2 (psaB) (P51-B).
 GN PSAB.
 OS Antirrhinum majus (Garden snapdragon).
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Eudicotyledons; Core eudicots;
 OC Asteridae; euasterids I; Fumiales; Veroniceae; Antirrhinum.
 OX NCBI_TaxID=4151;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=SO; TISSUE=Leaf;
 RX MEDLINE=96133654; PubMed=8544819;
 RA Schaffner C., Laasch H., Hagemann R.;
 FT "Detection of point mutations in chloroplast genes of Antirrhinum
 majus L. I. Identification of a point mutation in the psaB gene of a
 photosystem I plastome mutant.";
 RL Mol. Gen. Genet. 249:533-544(1995).
 CC -!- FUNCTION: Psaa and psaB bind p700, the primary electron donor of
 CC photosystem I (PSI), as well as the electron acceptors A0, A1, and
 CC FX. PSI functions as a plastocyanin/cytochrome c6-ferredoxin
 CC oxidoreductase.
 CC -!- COPOLYMER: p700 is a chlorophyll A dimer, A0 is chlorophyll A, A1
 CC is a phylloquinone and FX is a 4Fe-4S iron-sulfur center.
 CC -!- SUBUNIT: A psaa/B heterodimer binds the p700 chlorophyll special
 CC pair and subsequent electron acceptors. The PSI reaction center of
 CC higher plants and algae is composed of one at least 11 subunits.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
 CC thylakoid membrane.
 CC -!- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X84153; CAA58958.1; -;
 DR Nendel; 13513; ANTMA.psaB.2.
 DR InterPro: IPR001280; Psaa.psaB.
 DR Pfam: PF00223; psaa.psaB; 1.
 DR PRINTS: PR00257; PHOTOSYPSAAH.
 DR PROSITE: PS00419; PHOTOSYSTEM_I_PSAH; 1.
 KW Photosynthesis; Photosystem I; Electron transport; Chloroplast;
 KW Thylakoid; Transmembrane; Iron-sulfur; 4Fe-4S; Chlorophyll.
 FT TRANSFM: 46 69
 FT TRANSFM: 135 158
 FT TRANSFM: 175 199
 FT TRANSFM: 273 291
 FT TRANSFM: 330 353
 FT TRANSFM: 369 395
 FT TRANSFM: 417 439
 FT TRANSFM: 517 535
 FT TRANSFM: 575 596
 FT TRANSFM: 643 665
 FT TRANSFM: 707 727

FI METAL 559 559 IRON-SULFUR EX (4FE-4S) (BY SIMILARITY).
 FT METAL 568 568 IRON-SULFUR EX (4FE-4S) (BY SIMILARITY).
 FT BINDING 654 654 P700 SPECIAL PAIR CHLOROPHYLL AXIAL LIGAND (BY SIMILARITY).
 FT BINDING 662 662 A0 CHLOROPHYLL (BY SIMILARITY).
 FT BINDING 670 670 A0 CHLOROPHYLL (BY SIMILARITY).
 FT BINDING 671 671 A1 PHYLLOQUINONE (BY SIMILARITY).
 FT BINDING 700 700 A1 PHYLLOQUINONE (BY SIMILARITY).
 SQ SEQUENCE 734 AA: 82420 MW; AR3289601255430 / CRC64;

 Query Match 73.6%; Score 39; DB 1; Length 734;
 Best Local Similarity 75.0%; Pred. No. 9;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

 QY 2 QGNFNAWV 9
 |||||
 Db 71 QGNFESWV 78

 RESULT 10
 PSAB_SPIOL STANDARD; PRT; 734 AA.
 AC P06512;
 DT 01-JAN-1988 (rel. 06, Created)
 DT 01-JAN-1988 (rel. 06, Last sequence update)
 DT 01-MAR-2002 (rel. 41, Last annotation update)
 DE Photosystem I P700 chlorophyll A apoprotein A2 (psab) (181 B).
 GN PSAB.
 OS Spinacia oleracea (Spinach).
 OG Eukaryota.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Caryophyllales; Caryophyllales; Chenopodiaceae; Spinacia.
 OX NCBI_taxid=3562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kirsch W., Seyer P., Herrmann R.G.;
 RT "Nucleotide sequence of the clustered genes for two P700 chlorophyll
 a apoproteins of the photosystem I reaction center and the ribosomal
 protein S14 of the spinach plastid chromosome.";
 RL Curr. Genet. 10:843-855(1986).
 RZ [2]
 RP SEQUENCE FROM N.A.
 PC STPAIN-CV GRANT G/HIVEP, AND CV MORAPPO;
 PX MEDLINE=21187424; PubMed=11392074;
 FA Schmitz-Liknewer C., Maiti F.M., Albert J.F., Gattet A.,
 RA Herrmann R.G., Macho P.;
 RT "The plastid chromosome of spinach (Spinacia oleracea): complete
 nucleotide sequence and gene organization.";
 RL Plant Mol. Biol. 45:307-315(2001).
 CC -1- FUNCTION: psaa and psab bind P700, the primary electron donor of
 CC photosystem I (PSI), as well as the electron acceptors A0, A1, and
 CC FX. PSI functions as a plastocyanin/cytochrome c6-ferredoxin
 CC oxidoreductase.
 CC -1- COFACTOR: pmh is a chlorophyll A dimer. A0 is chlorophyll A. A1
 CC is a phytylquinone and FX is a 4Fe-4S iron-sulfur center.
 CC -1- SUPPLEMENT: A psaa/b heterodimer binds the P700 light-harvesting
 CC pair and subsequent electron acceptors. The PSI reaction center of
 CC higher plants and algae is composed of one at least 11 subunits.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast.
 CC Thylakoid membrane.
 CC -1- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its distribution is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.sib.ch/ftp/pub/seq/>
 CC or send an email to license@sib.ch.)
 CC
 CC EMBL: A140084; CAB88725.1.

DR PIR; S00445; S00445.
 DR Mendel; 2573; SPIOL; psab; 1.
 DR InterPro: IPR001282; psaa; psab.
 DR Pfam: PF00223; psaa; psab; 1.
 DR PRINTS: PR00257; PHOTOSYPSAAB.
 DR PROSITE: PS00419; PHOTOSYSTEM_I_PSAAB; 1.
 DE Photosynthesis; Photosystem I; Electron transport; Chloroplast;
 DE Thylakoid; Transmembrane; Iron-sulfur; 4Fe-4S; Chlorophyll.
 FT TRANSMEM 46 49
 FT TRANSMEM 135 158
 FT TRANSMEM 175 199
 FT TRANSMEM 273 291
 FT TRANSMEM 330 353
 FT TRANSMEM 369 395
 FT TRANSMEM 417 439
 FT TRANSMEM 517 535
 FT TRANSMEM 575 596
 FT TRANSMEM 643 665
 FT TRANSMEM 707 727
 FT METAL 559 559
 FT METAL 568 568
 FT BINDING 654 654
 FT BINDING 662 662
 FT BINDING 670 670
 FT BINDING 671 671
 FT BINDING 700 700
 SQ SEQUENCE 734 AA: 82129 MW; 9C1883B03348C22F CRC64;

 Query Match 73.6%; Score 39; DB 1; Length 734;
 Best Local Similarity 75.0%; Pred. No. 9;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

 QY 2 QGNFNAWV 9
 |||||
 Db 71 QGNFESWV 78

 RESULT 11
 PSAB_TOBAC STANDARD; PRT; 734 AA.
 AC P06407;
 DT 01-JAN-1988 (rel. 06, Created)
 DT 01-JAN-1988 (rel. 06, Last sequence update)
 DT 01-MAR-2002 (rel. 41, Last annotation update)
 DE Photosystem I P700 chlorophyll A apoprotein A2 (psab) (PSI-B).
 GN PSAB.
 OS Nicotiana tabacum (Common tobacco).
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 OX NCBI_taxid=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV BRIGHT YELLOW 4;
 SA Zito G., F. Zito M., Zito M., Wikus J.T., Hayashida N.,
 SA Matsubayashi T., Zaita N., Chonwongse J., Okukata T.,
 SA Yamauchi-Shinozaki K., Ohto C., Torazawa K., Mero B.Y., Suda M.,
 SA Ikeno H., Kamehashira T., Yamada K., Kusuda J., Iakawa F., Kato A.,
 SA Toidoh N., Shimada H., Sudira M.;
 RT "The complete nucleotide sequence of the tobacco chloroplast genome:
 RT its gene organization and expression.";
 RL EMBL J. 5:2043-2049(1986).
 CC -1- FUNCTION: psaa and psab bind P700, the primary electron donor of
 CC photosystem I (PSI), as well as the electron acceptors A0, A1, and
 CC FX. PSI functions as a plastocyanin/cytochrome c6-ferredoxin
 CC oxidoreductase.
 CC -1- SUPPLEMENT: A psaa/b heterodimer binds the P700 chlorophyll A. A1
 CC is a phytylquinone and FX is a 4Fe-4S iron-sulfur center.
 CC -1- SUBUNIT: A psaa/b heterodimer binds the P700 chlorophyll special
 CC pair and subsequent electron acceptors. The PSI reaction center of

CC higher plants and algae is composed of one at least 13 subunits.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein; Chloroplast
 CC THYLAKOID MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: Z00044; CAA73551.1;
 DR PIR: A03466; A2NTP7.
 DR Mordol: 6442; N07n-psak.1
 DR InterPro: IPR001280; Psaa-psab.
 DR Pfam: pf00224; psaa-psab.1.
 DR PRINTS: PR00257; PH01SVSPSAB.
 DR PROSITE: PS00419; PHOTOSYSTEM_I_PSAAB; 1.
 KW Photosynthesis; Photosystem I; Electron transport; Chloroplast;
 KW Thylakoid; Transmembrane; Iron-sulfur; 4Fe-4S; Chlorophyll.
 FT TRANSMEM 46 69
 FT TRANSMEM 135 158
 FT TRANSMEM 175 199
 FT TRANSMEM 273 291
 FT TRANSMEM 330 353
 FT TRANSMEM 369 395
 FT TRANSMEM 417 439
 FT TRANSMEM 517 535
 FT TRANSMEM 575 596
 FT TRANSMEM 643 665
 FT TRANSMEM 707 727
 FT METAL 559 559
 FT METAL 568 568
 FT BINDING 654 654
 FT BINDING 662 662
 FT BINDING 670 670
 FT BINDING 671 671
 FT BINDING 700 700
 SQ SEQUENCE: 744 AA; 82466 MW; 3240AARef+414-Per CP964;

Query Match 73.6%; Score 39; DB 1; Length 744;
 Best Local Similarity 75.0%; Pred. No. 9;
 Matches 6; Conservation 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QGNFNAW 9
 IIII I I
 DB 71 QGNFESW 78

RESULT 12
 MFGM_PIG STANDARD: PRT: 409 AA.
 AC P79385;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 10-MAY-2000 (Rel. 39, Last annotation update)
 DE Lactadherin (Milk fat globule-EGF factor 8) (MFG-E8) (Sperm
 DE surface protein SP47) (P847).
 GN MFG8.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE: Testis;
 RA Enslin M.A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/Joint databases.

CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-

CC BINDING PROTEIN.
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
 CC SPERMATOZOAN.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sib-sib.ch).

DR EMBL: Y11684; CAA/2379.1;
 DR HSSP: P00740; IEDM.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000421; FA5B_C.
 DR Pfam: PF00008; EGF_2.
 DR Pfam: PF00754; F5_F8_Type_C_2.
 DR SMART: SM00181; EGF_2.
 DR SMART: SM00231; FA5B_C_2.
 DR PROSITE: PS00022; EGF_1; 2.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS01285; FA5B_C_1; 2.
 DR PROSITE: PS01286; FA5B_C_2; 2.
 KW Glycoprotein; Repeat; EGF-like domain.
 FT DOMAIN 2 41
 FT DOMAIN 44 88
 FT DOMAIN 91 247
 FT DOMAIN 252 409
 FT SITE 67 69
 FT SITE 67 69
 FT DISULFID 6 17
 FT DISULFID 11 29
 FT DISULFID 31 40
 FT DISULFID 91 247
 FT DISULFID 234 238
 FT DISULFID 252 409
 FT CARBOHYD 41 41
 FT CARBOHYD 372 372
 SQ SEQUENCE: 409 AA; 45325 MW; 4007AARef+29927A CP964;

Query Match 71.7%; Score 38; DB 1; Length 409;
 Best Local Similarity 85.7%; Pred. No. 7.7;
 Matches 6; Conservation 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QGNFNAW 8
 IIII I I
 DB 294 QGNFNAW 300

RESULT 13
 MFGM_BOVIN STANDARD: PRT: 427 AA.
 AC Q95114; Q27959; P79344;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8)
 DE (MGP57/53) (PAS-6/PAS-7 glycoprotein) (MFGM) (Sperm surface protein
 DE SP47) (P847) (Components 15/16).
 GN MFG8.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A. PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
 RC SUPAIN-HUTSLIN; TISSUE: Mammary gland;
 RA MEDLINE=97008954; PubMed=8856064;

PA Hvarreard L., Andersen M.H., Berglund L., Rasmussen J.T.,
 RA Petersen T.E.;
 RT "Characterization of glycoprotein PAS-6/7 from membranes of bovine
 RL milk fat globules.";
 RN Eur J Biochem 240:628-636(1996).
 PP [2]
 PP SEQUENCE OF 18-427 FROM N.A.
 PP TISSUE-Mammary gland;
 PX MEDLINE=96125736; PubMed=8541316;
 RA Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura K.,
 RA Matsuda T.;
 RT "Molecular cloning of glycoprotein antigens MGP57/53 recognized by
 RT monoclonal antibodies raised against bovine milk fat globule
 RT membrane.";
 RL Biochim. Biophys. Acta 1245:385-391(1995).
 PN [3]
 PN SEQUENCE OF 19-427 FROM N.A.
 PP TISSUE=Testis;
 PC Esslin M.A.;
 RA Submitted (NOV-1997) to the EMBL/Genbank/DBET databases.
 RN [4]
 PP SEQUENCE OF 149-146: 174-187; 233-246 AND 422-427.
 RC TISSUE=Milk;
 RX MEDLINE=93050576; PubMed=9485470;
 RA Mather I.H., Rangbhart L.R., Lane W.S.;
 RT "The major fat-globule membrane proteins, bovine components 15/16 and
 RT quinea-pig GP 55, are homologous to MGF-E8, a murine glycoprotein
 RT containing epidermal growth factor-like and factor v/vIII-like
 RT sequences.";
 RL Biochem. Mol. Biol. Int 29:545-554(1993)
 CC -1- FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF
 CC MAMMARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-
 CC BINDING PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PROMOTED BY ALTERNATIVE SPLICING. THE SHORT FORM
 CC LACKS 53 AMINO ACIDS WITHIN THE F5/8 TYPE C 1 DOMAIN
 CC -1- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN
 CC -1- PTM: THE 2 GLYCOPROTEIN GLYCANS CONSIST OF GAL, GLUCNAc AND FUC. WITH
 CC PROBABLY FUC AS REDUCING TERMINAL SUGAR.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.sib.ch/permissions/>
 CC or send an email to license@isb.sib.ch).
 CC
 DR EMBL: X91895; CAA62997.1; 1;
 DR EMBL: S80643; AAB35894.2; 1;
 DR FMBL: Y11719; CAA72406.1; 1;
 DR BSSP: p00740; 1PDM
 DR InterPro: IP0000561; EGF-like
 DR InterPro: IP000421; FACH_1
 DR Pfam: PF00008; EGF_2;
 DR Pfam: PF00054; F5/8_Type_C_2;
 DR SMART: SM00181; EGF_2;
 DR SMART: SM00231; F5/8C_2;
 DR PROSITE: PS00022; EGF_1;
 DR PROSITE: PS01186; EGF_2;
 DR PROSITE: PS01285; F5/8C_1;
 DR PROSITE: PS01286; F5/8C_2;
 DR Signal: Glycoprotein; Milk; repeat, EGF-like domain,
 KW Alternative splicing
 FT SIGNAL 1 18
 FT CHAIN 19 427 LACTADHERIN.
 FT DOMAIN 20 59 EGF-LIKE 1.
 FT DOMAIN 62 106 EGF-LIKE 2.
 FT DOMAIN 109 265 F5/8 TYPE C 1
 FT DOMAIN 272 427 F5/8 TYPE C 2
 FT SITE 85 87 CELL ATTACHMENT SITE (POTENTIAL).

FT DISULFID 24 45 BY SIMILARITY.
 FT DISULFID 29 47 BY SIMILARITY.
 FT DISULFID 49 58 BY SIMILARITY.
 FT DISULFID 66 77 BY SIMILARITY.
 FT DISULFID 71 94 BY SIMILARITY.
 FT DISULFID 96 105 BY SIMILARITY.
 FT DISULFID 109 265
 FT DISULFID 252 256
 FT DISULFID 270 427
 FT CARBOHYD 27 27 O-LINKED (FUC...) (IN PAS-6).
 FT CARBOHYD 34 34 N-LINKED (GLUCNAc...) (IN PAS-6
 FT CARBOHYD 59 55 AND PAS-7).
 FT CARBOHYD 227 227 N-LINKED (GLUCNAc...) (HIGH MANNOSE) (IN
 FT CARBOHYD 159 221 PAS-6).
 FT CARBOHYD 19 19 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 28 28 A > F (IN REF. 1).
 FT CONFLICT 28 28 L > O (IN REF. 1).
 SQ SEQUENCE 427 AA; 47411 MW; 4CHREB3A1DC4EB24 CRC64;
 Query Match: 71.7%; Score 39; Pos 1; Length 427;
 Best local similarity 85.7%; Pos. No. 8;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 QGNFNAW 8
 DB 312 QGNFNAW 318
 RESULT 14
 PSAB_AKATH STANDARD; PRT; 734 AA.
 ID PSAB_ARATH
 AC P55757;
 DI 30-MAY-2000 (Feb. 39, Created)
 DT 30-MAY-2000 (Feb. 39, Last sequence update)
 DT 01-MAY-2002 (Feb. 41, Last annotation update)
 LE 1000; SLEN 1 1709 chlorophyll A ap-protein A2 (PSAB) (PSI-B).
 GN PSAB
 OS Arabidopsis thaliana (Mouse-ear cress)
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; Malvales; Rosales;
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 CC NCBI_TaxID=3702;
 RN [1]
 RL SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20039611; PubMed=10574454;
 RA Sato S., Nakamura Y., Kaneko T., Asanizu E., Tabata S.;
 RT "Complete structure of the chloroplast genome of Arabidopsis
 RT thaliana.";
 RL DNA Res. 6:283-290(1999).
 CC -1- FUNCTION: PSAB and psap bind P700, the primary electron donor of
 CC photosystem I (PSI), as well as the electron acceptors A0, A1, and
 CC P1. PSI functions as a plastocyanin/viologenone de-ferredoxin
 CC oxidoreductase.
 CC -1- CCFAC708, P700 is a chlorophyll A dimer, A0 is chlorophyll A, A1
 CC is a phytylquinone and P1 is a 4Fe 4S iron sulfur center.
 CC -1- SUBUNIT: A psaa/B heterodimer binds the P700 chlorophyll special
 CC pair and a special electron acceptors. The PSI reaction center of
 CC higher plants and algae is composed of one at least 11 subunits.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
 CC thylakoid membrane.
 CC -1- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.sib.ch/permissions/>
 CC or send an email to license@isb.sib.ch).

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:23:15, Search time 122.86 Seconds
(without alignments)
12,673 Million cell updates/sec

Title: us-09-744-804-35
Perfect score: 53
Sequences: 1 KOGNENAW 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 56222 seqs, 1294729 residues

Total number of hits satisfying chosen parameters: 56222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0.1

Maximum Match 100%

Listing first 45 summaries

Database:

SPTREMBL_19:

- 1: sp_archaea:
- 2: sp_bacteria:
- 3: sp_fungi:
- 4: sp_human:
- 5: sp_invertebrate:
- 6: sp_mammal:
- 7: sp_mhc:
- 8: sp_organelle:
- 9: sp_phase:
- 10: sp_plant:
- 11: sp_rodent:
- 12: sp_virus:
- 13: sp_vertebrate:
- 14: sp_unclassified:
- 15: sp_virus:
- 16: sp_bacteriup:
- 17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	100.0	335	4	Q9BTL9
2	43	81.1	363	6	Q77718
3	42	79.2	282	8	Q9TRK3
4	42	79.2	591	9	Q9THJ9
5	42	79.2	653	7	Q9P38F
6	39	73.6	135	8	Q33331
7	38	71.7	94	4	P77085
8	38	71.7	291	8	Q9T8L4
9	38	71.7	301	8	Q9TRK4
10	38	71.7	544	8	Q9THJ4
11	38	71.7	604	8	Q9THJ4
12	37	69.8	297	10	Q943V6
13	37	69.8	345	10	Q943V6
14	37	69.8	379	10	Q941C7
15	37	69.8	502	4	Q97479
16	37	69.8	817	16	P73619

17	37	69.8	2432	12	Q91QP4
18	37	69.8	2433	12	Q91QP4
19	36	67.9	176	11	Q9CP38
20	36	67.9	135	15	Q9EP38
21	36	67.9	135	15	Q99E51
22	36	67.9	145	15	Q99E19
23	36	67.9	135	15	Q99DX1
24	36	67.9	257	2	Q98E21
25	36	67.9	295	5	Q909A3
26	36	67.9	345	2	Q94774
27	36	67.9	359	11	Q9JH8
28	36	67.9	359	11	Q9JH8
29	36	67.9	359	11	Q9JH8
30	36	67.9	359	11	Q9JH8
31	36	67.9	371	4	Q9HC73
32	36	67.9	480	4	Q43854
33	36	67.9	480	11	Q35474
34	36	67.9	482	5	Q9XV78
35	36	67.9	754	17	Q9F139
36	36	67.9	928	11	Q99872
37	36	67.9	1564	4	Q15877
38	35	66.0	81	15	Q91V89
39	35	66.0	135	15	Q99E91
40	35	66.0	147	2	Q97396
41	35	66.0	147	2	Q97351
42	35	66.0	147	2	Q94055
43	35	66.0	147	2	Q94043
44	35	66.0	147	2	Q94045
45	35	66.0	147	2	Q94047

ALIGNMENTS

RESULT 1

Q9BTL9 PRELIMINARY: PRT: 335 AA.
AC Q9BTL9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
DE SIMILAR TO MILK FAT GLOBULE BCF FACTOR B PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
PA Strassberg P.
PL Submitted (1997-05-11) to the EMBL/Genbank/DDBJ databases.
CC -!- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.
DR EMBL: BC003610; AAH03610.1; -.
DR HSSP: P08709; 1HF9.
DR InterPro: IPR000561; BCF-like.
DR InterPro: IPR001438; BCF-1.
DR InterPro: IPR000421; FA5B_C.
DR PRINTS: PR00010; EGFBL00D.
DR SMART: SM00181; EGF; 1.
DR SMART: SM00301; EGF-like; 1.
DR SMART: SM00231; FA5B_C; 2.
DR PROSITE: PS00222; EGF-1; ENEFNMN_1.
DR PROSITE: PS01186; EGF-2; 1.
DR PROSITE: PS01285; FA5B_C; 1.
DR PROSITE: PS01286; FA5B_C; 2.
KW EGF-like domain; Glycoprotein.
SQ SPQSNCP 335 AA; 37523 MW; Q9B4E5DB789E25 CRC64;

Query Match 100.0%; Score 53; DR 4; Length 335;
Best Local Similarity 100.0%; Pred No. 0.091;
Matches 4; Conservative 0; Mismatch 0; Indels 0; Gaps 0;

```

QY 1 KGNFNW 9
DB 271 KGNFNW 274

RESULT 2
Q97718 1
ID Q97718 PRELIMINARY; PRT: 263 AA
AC Q97718;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DI 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SPERM-MEMBRANE ASSOCIATED PROTEIN P47 (FRAGMENT).
OS Equus caballus (Horse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Grotzfel M., Tonfor-Peterson E.;
KL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases
DR EMBL: AJ010121; CAA09010.1; -;
DR HSSP: P00740; 1EDM;
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000421; FA58_C.
DR Pfam: PF00754; EGF_1;
DR Pfam: PF00754; FGF_type_C; 2
DR SMART: SM00181; EGF_1;
DR SMART: SM00231; FA58C; 2.
DR PROSITE: PS00922; EGF_1; UNPFWN_2.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS01285; FA58C_1; 1.
KW EGF like domain; glycoprotein.
FT NON_TER 1
FT NON_TER 363
SQ SEQUENCE 363 AA: 40744 MW: 1PR6495AFQ33RD CRC64.

Query Match 81.1%; Score 43; DB 6; Length 363.
Best Local Similarity 97.5%; Pred. No. 6.5;
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 1 KGNFNW 8
DB 254 KGNFNW 261

RESULT 3
Q9TNJ9 1
ID Q9TNJ9 PRELIMINARY; PRT: 280 AA.
AC Q9TNJ9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DI 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PHOTOSYSTEM I P700 APOPROTEIN A2 (FRAGMENT).
OS PSAB.
OC Anthoceros punctatus.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Anthocerotophyta.
OX NCBI_TaxID=3234;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=CALLUS;
RA Nishiyama T., Kato M.;
KL Molecular phylogenetic analysis among bryophytes and tracheophytes
KL based on combined data of plastid coded genes and the 18S rRNA gene.
KL Mol. Biol. Evol. 16:1027-1036(1999).
DR EMBL: AB01665; BAA83441.1; -;
DR HSSP: P25896; 1JB0.
DR InterPro: IPR001280; PSAB_PsAB.
RS SEQUENCE FROM N.A.
RC TISSUE=CALLUS;

Query Match 79.2%; Score 42; DB 8; Length 591;
Best Local Similarity 87.5%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGNFNW 9
DB 71 QGNFNW 78

RESULT 5
Q9RDB5 1
ID Q9RDB5 PRELIMINARY; PRT: 653 AA.
AC Q9RDB5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DI 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PUTATIVE GLYCOSYL TRANSFERASE.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN A3(2);

```

```

DR Pfam: PF00223; psaa_psaB; 1.
DR PRINTS: PF00257; PHOTOSYPSAAB.
KW Chloroplast.
FT NON_TER 282
SQ SEQUENCE 282 AA: 31691 MW: 14036A58508B6044 CRC64;

Query Match 79.2%; Score 42; DB 8; Length 282;
Best Local Similarity 87.5%; Pred. No. 7.5;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGNFNW 9
DB 71 QGNFNW 78

RESULT 4
Q9TNJ9 1
ID Q9TNJ9 PRELIMINARY; PRT: 591 AA.
AC Q9TNJ9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DI 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PHOTOSYSTEM I P700 APOPROTEIN A2 (FRAGMENT).
OS PSAB.
OC Sphagnum fallax.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
OC Sphagnopsida; Sphagnales; Sphagnaceae; Sphagnum.
OX NCBI_TaxID=53036;
RN [1]
RS SEQUENCE FROM N.A.
RC MEDLINE=9904147; PubMed=10474899;
RA Nishiyama T., Kato M.;
KL Molecular phylogenetic analysis among bryophytes and tracheophytes
KL based on combined data of plastid coded genes and the 18S rRNA gene.
KL Mol. Biol. Evol. 16:1027-1036(1999).
DR EMBL: AB013671; BAA83447.1; -;
DR HSSP: P25896; 1JB0.
DR InterPro: IPR001280; PSAA_PsAB.
RS SEQUENCE FROM N.A.
RC TISSUE=CALLUS;
DR PRINTS: PF00223; psaa_psaB; 1.
DR PROSITE: PS00419; PHOTOSYSTEM_I_PSAAB; 1.
KW Chloroplast.
FT NON_TER 591
SQ SEQUENCE 591 AA: 65591 MW: 3715F744A52D1442 CRC64;

Query Match 79.2%; Score 42; DB 8; Length 591;
Best Local Similarity 87.5%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGNFNW 9
DB 71 QGNFNW 78

RESULT 5
Q9RDB5 1
ID Q9RDB5 PRELIMINARY; PRT: 653 AA.
AC Q9RDB5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DI 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PUTATIVE GLYCOSYL TRANSFERASE.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN A3(2);

```


RA Seeger K.J., Harris D.,
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Thomson N.R., Parkhill J., Barrell R.G., Razaquar M.A.,
 RL Submitted (JAN 2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RX MEDLINE-97000351; PubMed-8843436;
 RA Rederbach M., Kiesel H.M., Penaparte D., Pichner A., Collum J.,
 RA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmid and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 FL Mol. Microbiol. 21:47-56(1995).
 DR EMBL; AL136058; CAB65566.1; .
 DR InterPro; IPR001173; Glycos_transf_2.
 DR Pfam; PF00535; Glycos_transf_2.1
 KW Transferase.
 SQ SEQUENCE 653 AA: 73150 MW: 97929848.581AAC CQC64;

Query Match 79.2%; Score 42; DB 2; Length 653;
 Best Local Similarity 66.7%; Pred. No. 19;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QGNFNAAW 9
 III:III
 DB 239 KGNYNAAW 247

RESULT 6
 Q3331
 ID Q3331 PRELIMINARY; PRT: 135 AA.
 DT 01-NOV-1996 (TREMREL. 01, Created)
 DI 01-NOV-1996 (TREMREL. 01, Last sequence update)
 DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
 DE PHOTOSYSTEM I SUBUNIT.
 GN psaB.
 OS Antirrhinum majus (Garden snapdragon).
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicotyledons;
 OC Asteridae; euasterids I; Lamiales; Veronicaceae; Antirrhinum.
 OX NCBI_TaxID=4151;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-LEAF;
 RX MEDLINE-96133694; PubMed-8544819;
 RA Schaffner C., Laasch H., Haqemann P.;
 RT "Detection of point mutations in chloroplast genes of Antirrhinum
 majus L. 1. Identification of a point mutation in the psaB gene of a
 photosystem I plastome mutant.";
 FL Mol. Gen. Genet. 246:533-544(1995).
 DR EMBL; X84152; CAA58957.1;
 DR HSP; P25896; LJB0.
 DR InterPro; IPR001280; PsaA_psaB.
 DR Pfam; PF00223; PsaA_psaB; 1.
 KW Chloroplast.
 SQ SEQUENCE 135 AA: 15530 MW: 90789996.791A92 CQC64;

Query Match 73.6%; Score 30; DB 8; Length 135;
 Best Local Similarity 75.0%; Pred. No. 12;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGNFNAAW 9
 III:III
 DB 71 QGNFNAAW 78

RESULT 7
 P77085
 ID P77085 PRELIMINARY; PRT: 64 AA.
 AC P77085;
 DT 01-FEB-1997 (TREMREL. 02, Created)
 DI 01-FEB-1997 (TREMREL. 02, Last sequence update)
 DE "Sequence of minutes 4-25 of Escherichia coli.";
 DE HYPOTHETICAL PROTEIN.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chung E., Aliev E., Aradjo R., Aparicio A., Davis K., Duncan M.,
 RA Federspiel N., Hyman P., Kallman S., Kemp C., Kurd O., Low H., Lin D.,
 RA Namath A., Oetner P., Roberts D., Schramm S., Davis R.W.;
 RT "Sequence of minutes 4-25 of Escherichia coli.";
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U82598; AAB40755.1; .
 KW Hypothetical protein.
 SQ SEQUENCE 64 AA: 7520 MW: 7225B9E5F6B0B7 CQC64;

Query Match 71.7%; Score 38; DB 2; Length 64;
 Best Local Similarity 67.7%; Pred. No. 9.1;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 GNFNAW 9
 III:III
 DB 35 GNFSAW 41

RESULT 8
 Q9TNL4
 ID Q9TNL4 PRELIMINARY; PRT: 291 AA.
 AC Q9TNL4;
 DT 01-MAY-2000 (TREMREL. 13, Created)
 DI 01-MAY-2000 (TREMREL. 13, Last sequence update)
 DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
 DE PHOTOSYSTEM I P700 APOPROTEIN A2 (FRAGMENT).
 GN PSAB.
 OS Physcomitrella patens (Moss).
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Embryophyta; Embryophyta; Tracheophyta;
 OC Bryopsida; Funariaceae; Funariales; Funariaceae; Physcomitrella.
 OX NCBI_TaxID=3218;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-99404147; PubMed 10474899;
 RA Nishiyama T., Kato M.;
 RT "Molecular phylogenetic analysis among bryophytes and tracheophytes
 based on combined data of plastid coded genes and the 16S rRNA gene.";
 FL Mol. Biol. Evol. 16:1027-1036(1999).
 DR EMBL; AA013653; BAA83429.1; .
 DR HSP; P25896; LJB0.
 DR InterPro; IPR001280; PsaA_psaB.
 DR Pfam; PF00223; PsaA_psaB; 1.
 DE PFINTS; PF00257; PHOTOSYSAAB.
 KW Chloroplast.
 FT NON_TER 291 291
 SQ SEQUENCE 291 AA: 32606 MW: 89997179F1E01DB CQC64;

Query Match 71.7%; Score 38; DB 8; Length 291;
 Best Local Similarity 85.7%; Pred. No. 41;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGNFNAAW 8
 III:III
 DB 71 QGNFNAAW 77

```

RESULT 9
Q9TNK9 PRELIMINARY: PRT: 301 AA.
ID Q9TNK9
AC Q9TNK9
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
DE PHOTOSYSTEM I P700 APOPROTEIN A2 (FRAGMENT).
GN PSAB.
OS Coleochaete nitellarum.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Coleochaetaceae; Coleochaete.
OX NCBI_TaxID-78178;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN DTEX th1261;
RA Nishiyama T., Kato M.;
RX MEDLINE-99404147; PubMed-10474899;
RT "Molecular phylogenetic analysis among bryophytes and tracheophytes
based on combined data of plastid coded genes and the 18S rRNA gene.";
RL Mol. Biol. Evol. 16:1027-1036(1999)
DR EMBL: AB013659; BAAB445.1; -.
DR HSSP: P25896; 1JH0.
DR InterPro: IPR001280; Psaa_psaB.
DR Pfam: PF00223; Psaa_psaB; 1.
DR PRINTS: PR00257; PHOTOSYPSAAR.
KW Chloroplast.
FT NON_TER 301
SQ SEQUENCE 301 AA: 33636 MW: 2572620A7D09247E CRC64.

Query Match 71.7%; Score 48; DB 8; Length 301;
Best Local Similarity 85.7%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGNFNAW 8
DB 71 QGNFEAW 77

RESULT 10
Q9TNK9 PRELIMINARY: PRT: 586 AA.
ID Q9TNK9
AC Q9TNK9
DT 01-MAY-2000 (TRENBLREL. 13, Created)
DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
DE PHOTOSYSTEM I P700 APOPROTEIN A2 (FRAGMENT).
GN PSAB.
OS Adiantum capillus-veneris (Fern).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Eukaryophyta; Tracheophyta;
Filicophyta; Filicopsida; Filicales; Adiantaceae; Adiantum.
OX NCBI_TaxID-13818;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-99404147; PubMed-10474899;
RA Nishiyama T., Kato M.;
RX "Molecular phylogenetic analysis among bryophytes and tracheophytes
based on combined data of plastid coded genes and the 18S rRNA gene.";
RL Mol. Biol. Evol. 16:1027-1036(1999)
DR EMBL: AB013681; BAAB345.1; -.
DR HSSP: P25896; 1JH0.
DR InterPro: IPR001280; Psaa_psaB.
DR Pfam: PF00223; Psaa_psaB; 1.
DR PRINTS: PR00257; PHOTOSYPSAAR.
DR PROSITE: PS00419; PHOTOSYSTEM_I_PSAAR; 1
KW Chloroplast.
FT NON_TER 586
SQ SEQUENCE 586 AA: 64468 MW: BA0356869E7B68 CRC64;

Query Match 71.7%; Score 38; DB 8; Length 586;
Best Local Similarity 85.7%; Pred. No. 91;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGNFNAW 8
DB 93 QGNFEAW 99

RESULT 12
Q94BV6 PRELIMINARY: PRT: 297 AA.
ID Q94BV6
AC Q94BV6
DT 01-DEC-2001 (TRENBLREL. 19, Created)
DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)
DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
DE AT5G27210.T21H.120.
OS Arabidopsis thaliana (Mouse ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Eukaryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
Cucurbitales; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID-3702;
RN [1]
RP SEQUENCE FROM N.A.
RC Chou K., Chen H., Kim C.J., Kanesaka F., Meyers M.C., Banb J.,
Rowser L., Carninci T., Dale J.M., Gibson H.A., Goldsmith A.D.,
Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
Miranda M., Narusaka M., Nguyen M., Onodera G.S., Palm C.J.,
Pham P., Quach H., Saito T., Saito M., Saito M., Saito M.,
Tang C.C., Toriumi M., Yamada K., Yu G., Yu S., Shinzaki K.,
Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis cDNA clones.";

```

FL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases
 DR EMBL: AY039860; AAK6964.1; -
 SQ SEQUENCE 297 AA; 34133 MW; 37268000A00F348 CR264;

Query Match 69.8%; Score 37; DB 10; Length 297;
 Best Local Similarity 62.5%; Pred. No. 64;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGNFNWV 9
 |||||
 DB 156 QGNFSLWI 163

RESULT 13
 Q9M304 PRELIMINARY: PRT; 345 AA.
 AC Q9M304;
 DT 01-OCT-2000 (TrEMBLrel 15, Created)
 DT 01-OCT-2000 (TrEMBLrel 15, Last sequence update)
 DI 01-DEC-2001 (TrEMBLrel 19, Last annotation update)
 DE BETA 1-3 GLUCANASE.
 GN Glc.
 OS Vitis vinifera (Grape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;
 OC Vitis.
 OX NCBI_TaxID:29760;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN:CV, PINOT NOIR;
 RC Seibicke T., Buchholz G., Piegner A., Kaspermeyer H.H.;
 FT "Investigations of re-3 gene expression in grape."
 FL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
 DE EMBL: AJ277900; CAA9154.1; -
 DR HSP; P12257; I2A00.
 DR InterPro: IPR000490; Glyco_hydro_17.
 DE Pfam: PF00333; Glyco_hydro_17; 1.
 DE PROSITE: PS00587; GLYCOSYL_HYDROL_147; 1.
 SQ SEQUENCE 345 AA; 37456 MW; 890916021310985 CR264;

Query Match 69.8%; Score 37; DB 10; Length 345;
 Best Local Similarity 75.0%; Pred. No. 76;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 QGNFNWV 9
 |||||
 DB 100 QGNFSLWI 107

RESULT 14
 Q941C7 PRELIMINARY: PRT; 379 AA.
 AC Q941C7;
 DT 01-DEC-2001 (TrEMBLrel 19, Created)
 DT 01-DEC-2001 (TrEMBLrel 19, Last sequence update)
 DI 01-DEC-2001 (TrEMBLrel 19, Last annotation update)
 DE HYPOTHETICAL 42.7 KDA PROTEIN (FRAGMENT).
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID:3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kimura K., Nakada Y., Itoh Y.;
 RT "Partial Arabidopsis thaliana cDNA encoding a protein of unknown
 function."
 RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases
 DE EMBL: AB052682; BAB59127.1; -
 KW Hypothetical protein.
 FT NON_TER 1

FL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases

DR EMBL: AY039860; AAK6964.1; -
 SQ SEQUENCE 297 AA; 34133 MW; 37268000A00F348 CR264;

Query Match 69.8%; Score 37; DB 10; Length 297;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 NFNWV 9
 |||||
 DB 117 NFNWV 122

RESULT 15
 Q07979 PRELIMINARY: PRT; 502 AA.
 AC Q07979;
 DT 01-NOV-1996 (TrEMBLrel 01, Created)
 DT 01-NOV-1996 (TrEMBLrel 01, Last sequence update)
 DI 01-MAR-2001 (TrEMBLrel 16, Last annotation update)
 DE CHROMOSOME XII PEAKING FRAMF OFF Y1903W.
 GN Y1903W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae;
 OX NCBI_TaxID:4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Obermaier B., Piravandi E., Pinke M.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.
 DE [2]
 RP SEQUENCE FROM N.A.
 PA MIPS;
 FL Submitted (MAY 1996) to the EMBL/GenBank/DDBJ databases.
 DE EMBL: Z74235; CAA92557.1; -
 DR S00; S0004021; Y1903W.
 SQ SEQUENCE 502 AA; 57794 MW; 8F8D01184DAE1F6F CR264;

Query Match 69.8%; Score 37; DB 3; Length 502;
 Best Local Similarity 62.5%; Pred. No. 11602;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KGNFNWV 8
 |||||
 DB 283 KYGNVNSW 290

Search completed: September 5, 2002, 15:31:00
 Job time: 465 sec

GenCore version 4.5
Copyright (c) 1999 GenCorp Computer Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:27:33 Search time 58.86 seconds
(without alignments)
3,735 Million cell updates/sec

Title: US-09-744-804-36
Perfect score: 49
Sequence: 1 NLLPPMKWT 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /seqs/630-Data/5/100/AA/MR pep *

2: /seqs/630-Data/5/100/FA/MR pep *

3: /seqs/630-Data/2/100/AA/MR pep *

4: /seqs/630-Data/2/100/FA/MR pep *

5: /seqs/630-Data/2/100/FA/COMB pep *

6: /seqs/630-Data/2/100/FA/COMB pep *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	100.0	160	2	US-08-162-402B-10
2	49	100.0	187	2	US-08-162-402B-6
3	49	100.0	465	2	US-08-162-402B-8
4	32	65.3	22	3	US-08-940-095-163
5	32	65.3	22	3	US-08-940-093-163
6	32	65.3	22	2	US-08-940-096-163
7	32	65.3	22	4	US-08-940-096-163
8	32	65.3	22	4	US-08-940-096-163
9	32	65.3	1147	1	US-08-131-365B-38
10	32	65.3	1147	2	US-08-668-123-38
11	31	63.3	342	3	US-08-785-528-1
12	31	63.3	342	3	US-08-728-603-17
13	30	61.2	39	4	US-08-640-915A-74
14	30	61.2	40	1	US-08-318-193-6
15	30	61.2	40	6	US-08-327-17
16	30	61.2	54	1	US-08-214-770-9
17	30	61.2	54	5	PCT US95-02885-9
18	30	61.2	852	2	US-09-070-960-3
19	30	61.2	852	3	US-09-357-746-3
20	30	61.2	854	2	US-09-070-060-4
21	30	61.2	854	3	US-09-357-746-4
22	30	61.2	1091	1	US-07-695-564-3
23	30	61.2	1091	1	US-08-241-387-4
24	29	59.2	50	4	US-09-177-249-262
25	29	59.2	182	2	US-08-874-832-17
26	29	59.2	182	3	US-09-997-233-17
27	29	59.2	197	1	US-08-339-152A-21

28 29 59.2 197 2 US-08-027-999B-12 Sequence 12, Appl

29 29 59.2 197 2 US-08-489-276A-12 Sequence 12, Appl

30 29 59.2 219 2 US-08-874-932-8 Sequence 8, Appl

31 29 59.2 219 2 US-09-097-233-8 Sequence 8, Appl

32 29 59.2 225 2 US-08-637-755B-127 Sequence 127, App

33 29 59.2 225 3 US-08-871-355A-127 Sequence 127, App

34 29 59.2 225 4 US-09-201-945-127 Sequence 127, App

35 29 59.2 409 4 US-09-358-741-419 Sequence 449, App

36 29 59.2 711 3 US-08-945-599-2 Sequence 2, Appl

37 29 59.2 781 1 US-89-373-134D-2 Sequence 2, Appl

38 29 59.2 781 2 US-09-114-637-2 Sequence 2, Appl

39 29 59.2 3033 1 US-07-925-695-5 Sequence 5, Appl

40 29 59.2 3033 1 US-07-925-695-8 Sequence 8, Appl

41 29 59.2 3033 1 US-07-925-695-9 Sequence 9, Appl

42 29 59.2 3287 2 US-08-477-451-7 Sequence 7, Appl

43 28 57.1 36 4 US-09-183-786-41 Sequence 41, Appl

44 28 57.1 36 4 US-09-567-995-41 Sequence 41, Appl

45 28 57.1 110 1 US-08-111-939-24 Sequence 24, Appl

ALIGNMENTS

RESULT 1

US-08-162-402B-10

: Sequence 10, Application US/08162402B

: Patent No. 5972437

: GENERAL INFORMATION:

: APPLICANT: CERIAI, ROBERTO L.

: APPLICANT: PETERSON, JERRY A.

: ATTORNEY: JAF ALZA, LAVAL, J.

: INVENTOR: N. J. FAL; S. HUBAN MILE PAT

: NUMBER OF SEQUENCES: 29

: CORRESPONDENCE ADDRESS:

: ADDRESS: Freely, S.brook; 5 Joplowski

: STREET: 444 South Flower St., 19th Floor

: CITY: Los Angeles

: STATE: CA

: COUNTRY: USA

: ZIP: 90071

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Diskette

: COMPUTER: IBM Compatible

: OPERATING SYSTEM: DOS

: SOFTWARE: FASTSEQ for Windows Version 2.0

: CURRENT APPLICATION DATA: US-08-162-402B

: ATTORNEY'S NAME: US-08-162-402B

: FILING DATE: 03-DEC-1993

: CLASSIFICATION: 435

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER:

: FILING DATE:

: ATTORNEY/AGENT INFORMATION:

: NAME: Amzel, Viviana

: REGISTRATION NUMBER: 30,930

: REFERENCE/DOCKET NUMBER: 766 38215

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 213-622-7700

: TELEFAX: 213-489-4210

: INDEX:

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 160 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-162-402B-10

Query Match: 100.0%, Score 49, SE 2, Length 160;

Best Local Similarity: 100.0%, Prod. No. 3.98;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLLRRMMVT 9
DB 63 NLLRRMMVT 71

RESULT 2
US-08-162-402B-6
; Sequence 6, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANT, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LARocca, DAVID J.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
; TITLE OF INVENTION: GLOBULE (HMPG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 387 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-162-402B-6

Query Match 100.0%; Score 49; DB 2; Length 387;
Best Local Similarity 100.0%; Pred. No. 0 18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLLRRMMVT 9
DB 141 NLLRRMMVT 139

RESULT 3
US-08-162-402B-8
; Sequence 8, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANT, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LARocca, DAVID J.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT

; TITLE OF INVENTION: GLOBULE (HMPG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 465 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-162-402B-8

Query Match 100.0%; Score 49; DB 2; Length 465;
Best Local Similarity 100.0%; Pred. No. 0 22;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLLRRMMVT 9
DB 209 NLLRRMMVT 217

RESULT 4
US-08-940-095-163
; Sequence 163, Application US/08040095
; Patent No. 6004925
; GENERAL INFORMATION:
; APPLICANT: Dasseux, Jean-Louis
; APPLICANT: Sekul, Renate
; APPLICANT: Buttner, Klaus
; APPLICANT: Cornut, Isabelle
; APPLICANT: Metz, Gunter
; APPLICANT: Dufourcq, Jean
; TITLE OF INVENTION: APOLIPROTEIN A-1 AGONISTS
; TITLE OF INVENTION: AND THEIR USE TO TREAT DYSLIPIDEMIC DISORDERS
; NUMBER OF SEQUENCES: 258
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036 2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

? SOFTWARE: FastSEQ Version 2.0
? CURRENT APPLICATION DATA: US/08/040-095
? FILING DATE: 29-SEP-1997
? CLASSIFICATION: 514
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Coruzzi, Laura A
? REGISTRATION NUMBER: 30,742
? REFERENCE/WORK KEY NUMBER: 000196-0004-999
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 650-493-4935
? TELEFAX: 650-493-5556
? TELEX: 66141 PENNIE
? INFORMATION FOR SEQ ID NO: 163:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 22 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: No. 604925e
? US-08-940-095-163

Query Match 65.3% Score 32, DB 3, Length 22,
Best Local Similarity 71.4% Pred. No. 12;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY 1 NLLRRW 7
||| |
DB 8 NLLRLW 14

RESULT 5
US-08-940-095-163
? Sequence 163, Application US/08940093
? Patent No. 6037323
? GENERAL INFORMATION:
? APPLICANT: Dasseux, Jean-Louis
? APPLICANT: Sekul, Renate
? APPLICANT: Buttner, Klaus
? APPLICANT: Cornut, Isabelle
? APPLICANT: Metz, Gunther
? TITLE OF INVENTION: APOLIPOPROTEIN A-1 AGONISTS
? TITLE OF INVENTION: AND THEIR USE TO TREAT DYSLIPIDEMIC DISORDERS
? NUMBER OF SEQUENCES: 258
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Pennie & Edmonds LLP
? STREET: 1155 Avenue of the Americas
? CITY: New York
? STATE: NY
? COUNTRY: USA
? ZIP: 10036-2811
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FastSEQ Version 2.0
? CURRENT APPLICATION DATA: US/08/040-095
? FILING DATE: 29-SEP-1997
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Coruzzi, Laura A
? REGISTRATION NUMBER: 30,742
? REFERENCE/WORK KEY NUMBER: 000196-0006-999
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 650-493-4935

? TELEFAX: 650-493-5556
? TELEX: 66141 PENNIE
? INFORMATION FOR SEQ ID NO: 164:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 22 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: No. 6047324e
? US-08-940-093-164

Query Match 65.3% Score 32, DB 3, Length 22,
Best Local Similarity 71.4% Pred. No. 12;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY 1 NLLRRW 7
||| |
DB 8 NLLRLW 14

RESULT 6
US-08-940-096-163
? Sequence 163, Application US/08940096
? Patent No. 6046166
? GENERAL INFORMATION:
? APPLICANT: Dasseux, Jean-Louis
? APPLICANT: Sekul, Renate
? APPLICANT: Buttner, Klaus
? APPLICANT: Cornut, Isabelle
? APPLICANT: Metz, Gunther
? TITLE OF INVENTION: APOLIPOPROTEIN A-1 AGONISTS
? TITLE OF INVENTION: AND THEIR USE TO TREAT DYSLIPIDEMIC DISORDERS
? NUMBER OF SEQUENCES: 258
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Pennie & Edmonds LLP
? STREET: 1155 Avenue of the Americas
? CITY: New York
? STATE: NY
? COUNTRY: USA
? ZIP: 10036-2811
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FastSEQ Version 2.0
? CURRENT APPLICATION DATA: US/08/040-096
? FILING DATE: 29-SEP-1997
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Coruzzi, Laura A
? REGISTRATION NUMBER: 30,742
? REFERENCE/WORK KEY NUMBER: 000196-0005-999
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 650-493-4935
? TELEFAX: 650-493-5556
? TELEX: 66141 PENNIE
? INFORMATION FOR SEQ ID NO: 163:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 22 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: No. 6046166
? US-08-940-096-163

Query Match 65.3% Score 32, DB 3, Length 22;
Best Local Similarity 71.4% Pred. No. 12;

Matches 5; Conservation 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLLRRMW 7
||| | |
Db 8 NLLERLW 14

RESULT 7

US-09-465-719-163
; Sequence 163, Application US/09465719
; Patent No. 6265477

GENERAL INFORMATION:

APPLICANT: Dassou, Jean-Louis
APPLICANT: Sekul, Renate
APPLICANT: Buttner, Klaus
APPLICANT: Cornut, Isabelle
APPLICANT: Melz, Gunther
TITLE OF INVENTION: APOLIPOPROTEIN A-1 AGONISTS
AND THEIR USE TO TREAT DYSLIPIDEMIC DISORDERS

TITLE OF INVENTION: APOLIPOPROTEIN A-1 AGONISTS

TITLE OF INVENTION: AND THEIR USE TO TREAT DYSLIPIDEMIC DISORDERS

NUMBER OF SEQUENCES: 258

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10036-2811

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 2.0

CURRENT APPLICATION DATA: US/09465, 719

APPLICATION NUMBER: 08/940,093

FILING DATE: 29-SEP-1997

CLASSIFICATION: 65.48

PRIOR APPLICATION DATA: 08/940,093

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 009196-0004-999

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-493-4935

TELEFAX: 650-493-5556

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 163:

SEQUENCE CHARACTERISTICS:

LENGTH: 22 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 6265377e

US-09-465-719-163

Query Match

Best Local Similarity 71.4%; Score 32; DB 4; Length 22;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLLRRMW 7
||| | |
Db 8 NLLERLW 14

RESULT 8

US-09-465-605-163
; Sequence 163, Application US/09453605
; Patent No. 6329341

GENERAL INFORMATION:

APPLICANT: Dassou, Jean-Louis

APPLICANT: Sekul, Renate

Buttner, Klaus
Cornut, Isabelle
Melz, Gunther
Dufourcq, Jean

TITLE OF INVENTION: APOLIPOPROTEIN A-1 AGONISTS

AND THEIR USE TO TREAT DYSLIPIDEMIC DISORDERS

NUMBER OF SEQUENCES: 258

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: NY

COUNTRY: USA

ZIP: 10036-2811

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09465, 605

FILING DATE: 26 No. 6329341-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/940,095

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 009196-0004-999

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-493-4935

TELEFAX: 650-493-5556

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 163:

SEQUENCE CHARACTERISTICS:

LENGTH: 22 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: No. 6329341e

SEQUENCE DESCRIPTION: SEQ ID NO: 163;

US-09-453-605-163

Query Match

Best Local Similarity 71.4%; Score 32; DB 4; Length 22;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLLRRMW 7
||| | |
Db 8 NLLERLW 14

RESULT 9

US-08-131-365B-38

; Sequence 38, Application US/08131365B

; Patent No. 5527690

GENERAL INFORMATION:

APPLICANT: Brown, Michael S.

APPLICANT: Briggs, Michael R.

APPLICANT: Wang, Xiaodong

APPLICANT: Goldstein, Joseph L.

TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING

TO STEROL REGULATORY ELEMENT BINDING

TITLE OF INVENTION: PROTEINS

NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Burke

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: U.S.A.

ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/731,365K
FILING DATE: 01-OCT-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UTSD:372/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 1147 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-131-365B-38

Query Match 65.3%; Score 32; DB 1; Length 1147;
Best Local Similarity 62.5%; Pred. No. 50-02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Caps 0;

QY 1 NLLRPMWV 8
:||||:|
Db 634 HLLQRLWV 641

RESULT 10
US-08-668-123-38
Sequence 38, Application US/08668123
Patent No. 5891631
GENERAL INFORMATION:
APPLICANT: Brown, Michael S.
APPLICANT: Briggs, Michael R.
APPLICANT: Wang, Xiaodun
APPLICANT: Goldstein, Joseph L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
TO SKELETAL REGULATOR ELEMENT BINDING
TITLE OF INVENTION: PROTEINS
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: U.S.A.
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/668,123
FILING DATE: 14-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/141,465
FILING DATE: 01-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UTSD:372/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000

TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 1147 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-668-123-38

Query Match 65.3%; Score 32; DB 2; Length 1147;
Best Local Similarity 62.5%; Pred. No. 50-02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Caps 0;

QY 1 NLLRPMWV 8
:||||:|
Db 634 HLLQRLWV 641

RESULT 11
US-08-785-928-1
Sequence 1, Application US/08785928
Patent No. 6087115
GENERAL INFORMATION:
APPLICANT: Gershenjorn, Marvin C.
APPLICANT: Arvanitakis, Leandros
APPLICANT: Geras-Raaka, Elizabeth
APPLICANT: Caserman, Ethel
TITLE OF INVENTION: CONSTITUTIVELY ACTIVE G PROTEIN COUPLED
RECEPTOR OF HHV 8 AND METHOD OF IDENTIFYING NEGATIVE
TITLE OF INVENTION: ANTAGONISTS OF G PROTEIN COUPLED RECEPTORS
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Barbara, Evans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/785,928
FILING DATE: 22-JAN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/1420
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 342 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-785-928-1

Query Match 64.4%; Score 31; DB 3; Length 342;
Best Local Similarity 71.4%; Pred. No. 240-02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Caps 0;

QY 2 LLRPMWV 8
:||||:
Db 276 LLRPMWV 282

RESULT 12

US-08-728-603-17
 ? Sequence 17, Application US/08728603
 ? Patent No. 6093806
 ? GENERAL INFORMATION:
 ? APPLICANT: Cosarman, Ethel
 ? APPLICANT: Knowles, Daniel M.
 ? TITLE OF INVENTION: PROTEINS OF KAPOSI'S SARCOMA ASSOCIATED
 ? WITH HUMAN HERPESVIRUS
 ? NUMBER OF SEQUENCES: 19
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: NIXON, HAPPAVE, DEVANS & POYLE LLP
 ? STREET: Clinton Square, P.O. Box 1051
 ? CITY: Rochester
 ? STATE: New York
 ? COUNTRY: USA
 ? ZIP: 14603
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Floppy disk
 ? COMPUTER: IBM PC compatible
 ? OPERATING SYSTEM: PC-DOS/MS-DOS
 ? SOFTWARE: Patent In Release #1.0, Version #1.30
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/08/728,603
 ? FILING DATE: 10-OCT-1996
 ? CLASSIFICATION: 435
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: BRAMAN, SUSAN J.
 ? REGISTRATION NUMBER: 34,103
 ? REFERENCE/TITLE NUMBER: 3663,725
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: 716-263-1636
 ? TELEFAX: 716-263-1600
 ? INFORMATION FOR SEQ ID NO: 17:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 342 amino acids
 ? TYPE: amino acid
 ? STRANDEDNESS: linear
 ? TOPOLOGY: protein
 ? MOLECULE TYPE: protein
 ? US-08-728-603-17

Query Match 63.3% Score 317 DB 37 Length 342
 Best Local Similarity 71.4% Pred. No. 2,46,62
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLRRMW 8

DB 276 LLRRMW 282

RESULT 13

US-08-630-915A-73
 ? Sequence 73, Application US/08630915A
 ? Patent No. 6409820
 ? GENERAL INFORMATION:
 ? APPLICANT: SPARKS, Andrew B.
 ? APPLICANT: BOFFMAN, No. 6309820h
 ? APPLICANT: KAY, Brian K.
 ? APPLICANT: FOWLES, Dana M.
 ? TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
 ? TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
 ? NUMBER OF SEQUENCES: 227
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Domingo & Edwards LLP
 ? STREET: 1155 Avenue of the Americas
 ? CITY: New York
 ? STATE: New York
 ? COUNTRY: USA

ZIP: 10036-2711
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Floppy disk
 ? COMPUTER: IBM PC compatible
 ? OPERATING SYSTEM: PC-DOS/MS-DOS
 ? SOFTWARE: Patent In Release #1.0, Version #1.30
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/08/630,915A
 ? FILING DATE: 03-APR-1996
 ? CLASSIFICATION: 536
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: MISTOCK, S. Leslie
 ? REGISTRATION NUMBER: 18,872
 ? REFERENCE/TITLE NUMBER: 1101-174
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: (212) 790-9090
 ? TELEFAX: (212) 866-8864/9741
 ? TELEX: 66141 PENNIE
 ? INFORMATION FOR SEQ ID NO: 73:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 39 amino acids
 ? TYPE: amino acid
 ? STRANDEDNESS:
 ? TOPOLOGY: unknown
 ? MOLECULE TYPE: peptide
 ? US-08-630-915A-73

Query Match 61.2% Score 30, DB 4, Length 39;
 Best Local Similarity 66.7% Pred. No. 45;
 Matches 6; Conservative 0; Mismatches 0; Gaps 0;

QY 1 NLIRHMV 9

DB 27 NELEDMV 35

RESULT 14

US-08-318-193-6
 ? Sequence 6, Application US/08318193
 ? Patent No. 5641663
 ? GENERAL INFORMATION:
 ? APPLICANT: GARVIN, Robert T.
 ? TITLE OF INVENTION: AN EXPRESSION SYSTEM FOR THE SECRETION
 ? TITLE OF INVENTION: OF BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY
 ? TITLE OF INVENTION: STIMULATING FACTOR (GM-CSF) AND OTHER HETEROLOGOUS
 ? TITLE OF INVENTION: PROTEINS FROM STREPTOMYCES
 ? NUMBER OF SEQUENCES: 91
 ? CORRESPONDENCE ADDRESS:
 ? ADDRESSEE: Foley & Lardner
 ? STREET: 1400 Diagonal Road, Suite 500
 ? CITY: Alexandria
 ? STATE: Virginia
 ? COUNTRY: USA
 ? ZIP: 22313-0599
 ? COMPUTER READABLE FORM:
 ? MEDIUM TYPE: Floppy disk
 ? COMPUTER: IBM PC compatible
 ? OPERATING SYSTEM: PC-DOS/MS-DOS
 ? SOFTWARE: Patent In Release #1.0, Version #1.25
 ? CURRENT APPLICATION DATA:
 ? APPLICATION NUMBER: US/08/318,193
 ? FILING DATE:
 ? CLASSIFICATION: 435
 ? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: US/07/935,314
 ? FILING DATE:
 ? APPLICATION NUMBER: US/07/224,568
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: BENI, Stephen A.
 ? REGISTRATION NUMBER: 29,768
 ? REFERENCE/TITLE NUMBER: 18746/116 CACO

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9400
TELEFAX: (703)883-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-318-193-6

Query Match 61.2%; Score 30; DB 1; Length 40;
Best Local Similarity 44.4%; Pred. No. 46;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLLRRMWT 9
| | | | |
Db 11 NAARRVWIS 19

RESULT 15
5200327-17
PATENT NO. 5200327
APPLICANT: GARVIN, ROBERT J.; MALEK, LAWRENCE T.
TITLE OF INVENTION: EXPRESSION SYSTEM FOR THE SECRETION OF
BIOACTIVE HUMAN GRANULOCYTE MACROPHAGE COLONY STIMULATING
FACTOR (GM-CSF) AND OTHER HETEROLOGOUS PROTEINS FROM
STREPTOMYCES
NUMBER OF SEQUENCES: 24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/97/224,568
FILING DATE: 26-JUL-1988
SEQ ID NO: 17:
LENGTH: 40
5200327-17

Query Match 61.2%; Score 30; DB 6; Length 40;
Best Local Similarity 44.4%; Pred. No. 46;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLLRRMWT 9
| | | | |
Db 11 NAARRVWIS 19

Search completed: September 5, 2002, 15:27:34
Job time: 319 sec

GenCore version 4.5
Copyright (c) 1993 2000 CompuGen Ltd.

OM protein protein search, using sw model

Run on: September 5, 2002 15:31:39 : Search time 23.59 seconds
(without alignments)
10,252 Million cell updates/sec

Title: US-09-744-904-36
Perfect score: 49
Sequence: 1 NLRKMWVT 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	49	100.0	387	1 MFCM_HUMAN	Q08431 homo sapien
2	44	89.8	427	1 MFCM_BOVIN	Q95114 bos taurus
3	38	77.5	224	1 FLCA_YEREN	Q56892 yersinia co
4	35	71.4	115	1 PK20_MARPO	P06385 marcanthia
5	35	71.4	118	1 PK20_PINTH	P41610 pinus thunb
6	35	71.4	186	1 P152_METTH	Q56255 methanobact
7	35	71.4	409	1 MFCM_PIG	P76385 sus scrofa
8	35	71.4	909	1 ORO1_RUCAL	P57388 buchnera ap
9	34	69.4	111	1 P220_CHLPE	P36545 chlamydomon
10	34	69.4	699	1 NUSC_DIGGR	Q42131 digitalis g
11	34	69.4	1281	1 YLR5_CAFRI	P46580 caenorhabdi
12	33	67.3	234	1 ARGR_HUMAN	P55145 homo sapien
13	33	67.3	468	1 Y195_CAFRI	P49049 caenorhabdi
14	32	65.2	320	1 Y4P3_BRIGN	P55442 rhizobium s
15	32	65.3	354	1 GBI2_CHICK	P50147 gallus gall
16	32	65.3	360	1 YF45_HAFIN	P41246 haemophilus
17	32	65.3	344	1 VIF4_FPGAV	P43358 frog virus
18	32	65.3	445	1 EX7L_PASMU	Q9cp86 pasteurilla
19	32	65.3	582	1 PGMU_FOFTN	Q9cs94 populus tre
20	32	65.3	1924	1 SPF1_PAT	P56720 rattus norv
21	32	65.3	1952	1 Y170_GCMVA	P11419 human cytoch
22	32	65.3	1133	1 SRG1_F4_CSE	Q56816 erichsalus
23	32	65.3	1147	1 SRG1_HUMAN	P46956 homo sapien
24	32	65.3	1390	1 INSR_AEDAE	Q93105 aedes aegypt
25	32	65.3	3305	1 APUR_MANSE	Q25490 manduca sex
26	31	63.3	115	1 RK20_PCHPO	P41294 carphyrus fu
27	31	63.3	117	1 RK20_ARATH	P56794 arabidopsis
28	31	63.3	117	1 FL20_SYNY3	P48957 symbiodi
29	31	63.3	117	1 PK20_MAIFP	P56566 sea urchin
30	31	63.3	127	1 PK20_SF101	P56903 spirochaeta
31	31	63.3	127	1 PK20_TORAC	P06386 nicotiana gl
32	31	63.3	230	1 CPNR_STPA7	Q04462 streptococ
33	31	63.3	241	1 P13_AQUAE	Q56431 apulifox 900

RESULT 1

ID	MFCM_HUMAN	STANDARD	PRT	387 AA
AC	Q08431			
DT	01-OCT-1996	(Ref. 34, Created)		
DT	01-NOV-1997	(Ref. 35, Last sequence update)		
DT	16-OCT-2001	(Ref. 40, Last annotation update)		
DE	Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (HMFG)			
DE	(breast epithelial antigen BA45) (MFCM) [Contains: Madin].			
GN	MFG88			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
OX	NCBI_TaxID=9606;			
RP	SEQUENCE FROM N.A.			
RP	TISSUE=Breast, and Breast carcinoma;			
RX	MEDLINE 96213908; PubMed=8639264;			
RA	Couto J.R., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.,			
RA	cloning and sequence analysis of human breast epithelial antigen			
RT	BA45 reveals an EGF cell adhesion sequence presented on an epidermal			
RT	growth factor-like domain.";			
RL	DNA Cell Biol. 15:281-286(1996).			
RN	[2]			
RF	SEQUENCE OF 170-387 FROM N.A.			
RC	TISSUE=Mammary gland;			
PX	MEDLINE 91371351; PubMed=1909452;			
KA	Larocca D., Peterson J.A., Urrea R., Kusiyoshi J., Bistrian A.M.,			
RA	Ceriani R.L.;			
RT	"A Mr 46,000 human milk fat globulin protein that is highly expressed			
RT	in human breast tumors contains factor VIII-like domains.";			
RE	Cancer Res. 51:4994-4998(1991).			
RN	[3]			
RF	PARTIAL SEQUENCE, AND CHARACTERIZATION.			
RC	TISSUE=Milk;			
PX	MEDLINE 98194924; PubMed=9535276;			
KA	Guaffrida M.G., Guaffrida M., Giotta C., Conti A.,			
RA	Godard-Zimmermann J.;			
RT	"Isolation and characterization of full and truncated forms of human			
RT	breast carcinoma protein BA45 from human milk fat globule membranes.";			
RL	J. Protein Chem. 17:143-148(1998).			
RN	[4]			
RF	SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.			
PX	MEDLINE 9942876; PubMed 10411933;			
KA	Boqvist E., Rastad J., Fritton E., Westermarck J., Mucchiano G.,			
RA	Tjernberg L.G., Nordstedt C., Engstrom U., Westermarck P.;			
RT	"Medin, an integral component of aortic smooth muscle cell-produced			
RT	Lactadherin forms the most common human amyloid.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 93:8474(1996).			
RN	[5]			
RF	CHARACTERIZATION.			
PX	MEDLINE 97405885; PubMed 9269029;			
KA	Taylor M.F., Ceriani R.L., Ceriani R.L., Peterson J.A.;			
RA	"Lactadherin (formerly BA45), a membrane associated glycoprotein			
RT	expressed in human milk and breast carcinomas, promotes Arg Gly-Asp			
RT	(RGD)-dependent cell adhesion.";			

RL UNA cell Biol. 16:861-869(1997).
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
 CC SPECIFICALLY TO PHOSPHATIDYLCHOLINE AND INHIBITS ITS REPLICATION.
 CC -1- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF APOPTIC MEDIAL AMYLOID
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND APOPTIC
 CC MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
 CC -1- PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
 CC AMINO ACID 264 AND 273.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U58516; AAC50549.1; -;
 DR EMBL: S56151; AAL19771.1; -;
 DR EMBL: 642241.1; -;
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000421; FAS8_C.
 DR Pfam: PF00008; EGF_1.
 DR Pfam: PF00754; F5_F8_type_C; 2.
 DR SMART: SM00181; EGF_1.
 DR SMART: SM00231; FAS8C; 2.
 DR PROSITE: PS01285; EGF_1; 1.
 DR PROSITE: PS01285; FAS8C; 1; 2.
 DR PROSITE: PS01285; FAS8C; 2; 2.
 KW Signal; glycoprotein; Milk; Repeat; EGF-like domain; Amyloid.
 FT SIGNAL 1 23
 FT CHAIN 24 387 LACTADHERIN.
 FT CHAIN 202 387 LACTADHERIN: SHORT FORM
 FT CHAIN 268 317 MEDIN.
 FT DOMAIN 24 67 EGF-1;F5.
 FT DOMAIN 70 225 F5/8 TYPE C 1.
 FT DOMAIN 230 387 F5/8 TYPE C 2.
 FT SITE 46 48 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 27 38 BY SIMILARITY.
 FT DISULFID 32 55 BY SIMILARITY.
 FT DISULFID 57 66 BY SIMILARITY.
 FT DISULFID 70 225 BY SIMILARITY.
 FT DISULFID 212 216 BY SIMILARITY.
 FT DISULFID 230 387 BY SIMILARITY.
 FT CARBOHYD 238 238 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 325 325 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 329 329 N-LINKED (GLCNAC) (POTENTIAL).
 FT CARBOHYD 350 350 N-LINKED (GLCNAC) (POTENTIAL).
 SQ SEQUENCE 387 AA; 43123 MW; 2EE6571BDCB3/62D CRC64;
 Query Match: 100.0%; Score 49; DB 1; Length 387.
 Best Local Similarity: 100.0%; Pos. Neg. 0.021;
 Matches 9; Conservative 0; Mismatches 0; Gaps 0;
 QY 1 NLLRRMMWT 9
 DB 131 NLLRRMMWT 139
 RESULT 2
 ID MFGM_BOVIN STANDARD; PRT; 427 AA.
 AC Q95114; Q27459; P79344;
 DT 01-NOV-1997 (Ref. 35, Created)
 DT 01-NOV-1997 (Ref. 35, Last sequence update)
 DT 16-OCT-2001 (Ref. 40, Last annotation update)
 DE Lactadherin (Medin) (Milk fat globule EGF factor 2) (MFG2_EGF)
 DE (MGF57/53) (FAS-6/PAS-7 glycoprotein) (MFGM) (Sperm surface protein

DE SP47) (BP47) (Components 15/16).
 GN MFG2E8.
 OS Bos taurus (Bovine).
 CC Pokaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Bovinae; Bos.
 CC NCBI_TaxID:9913;
 RN [1]
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
 RC STRAIN=HOLSTEIN; TISSUE=Mammary gland;
 RX MEDLINE=97008954; PubMed=8856064;
 RA Hovgaard T., Anderson M.R., Berglund L., Rasmussen J.T.,
 RA Petersen T.E.;
 FT "Characterization of glycoprotein FAS-6/7 from membranes of bovine
 FT milk fat globules.";
 FT Eur. J. Biochem. 240:628-636(1996).
 RN [2]
 RP SEQUENCE OF 18-427 FROM N.A.
 RC TISSUE=Mammary gland;
 RX MEDLINE=96125736; PubMed=8541416;
 RA Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,
 RA Matsuda T.;
 FT "Molecular cloning of glycoprotein milk fat globule EGF-2/53 recognized by
 FT monoclonal antibodies raised against bovine milk fat globule
 FT membrane.";
 FT Biochim. Biophys. Acta 1245:385-391(1995).
 RN [3]
 RP SEQUENCE OF 19-427 FROM N.A.
 RC TISSUE=Testis;
 RA Ensslin M.A.;
 FT Published (N-V:557) in the EMBL, GenBank, DDBJ databases.
 RN [4]
 RP SEQUENCE OF 140-146; 174-187; 233-246 AND 422-427.
 RC TISSUE=Milk;
 RX MEDLINE=91250576; PubMed=8485470;
 RA Mather I.H., Banquhart L.R., Lane W.S.;
 FT "The major fat globule membrane proteins, bovine components 15/16 and
 FT guinea-pig GP 55, are homologous to MGF-E8, a murine glycoprotein
 FT containing epidermal growth factor-like and factor V/VIII-like
 FT sequences.";
 FT Biochem. Mol. Biol. Int. 29:545-554(1993).
 CC -1- FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF
 CC MAMMARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-
 CC BINDING PROTEIN.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM. ARE PRODUCED BY ALTERNATIVE SPLICING. THE SHORT FORM
 CC LACKS 53 AMINO ACIDS WITHIN THE F5/8 TYPE C 1 DOMAIN.
 CC -1- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN.
 CC -1- PTM: THE 2 N-LINKED GLYCANS CONSIST OF GAL, GLCNAC AND FUC, WITH
 CC PROBABLY FUC AS PEPTIC TERMINAL SUGAR.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X91895; CAA62997.1; -;
 DR EMBL: S80643; AAR35894.2; -;
 DR EMBL: Y11719; CAA72406.1; -;
 DR EMBL: P00740; ICDM
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000421; FAS8_C.
 DR Pfam: PF00008; EGF_1.
 DR Pfam: PF00754; F5_F8_type_C; 2.
 DR SMART: SM00181; EGF_1.
 DR SMART: SM00231; FAS8C; 2.
 DR PROSITE: PS01285; EGF_1; 1.
 DR PROSITE: PS01285; EGF_2; 2.
 DR PROSITE: PS01285; FAS8C; 2; 2.


```

DR PRINTS: P00062; RIBOSOMAL_L20.
DR Problem: P0002486; RIBOSOMAL_L20; 1.
DR PROSITE: P00037; RNA-binding; Chloroplast.
KW Ribosomal protein; RNA-binding; Chloroplast.
FT INIT_MET 0 BY SIMILARITY.
SQ SEQUENCE 115 AA; 13475 MW; A91B680F1D777A39 CRC64;

Query Match
Best Local Similarity 71.4%; Score 35; DB 1; Length 115;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LRRMMVT 9
DB 56 LRRIMVT 62
|||||

RESULT 5
RK20_PINTH STANDARD PRT: 119 AA
AC 141610;
DT 01-NOV-1995 (Rel. 42, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Chloroplast 50S ribosomal protein L20.
GN RPL20.
OS Pinus thunbergii (Green pine) (Japanese black pine).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta.
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus
OX NCBI_TaxID: 4350;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 95024047; PubMed: 7937893;
RA Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
RA Sugita M.;
RT "Loss of all adh genes as determined by sequencing the entire
RT chloroplast genome of the black pine Pinus thunbergii.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC -1- SIMILARITY: BELONGS TO THE L20P FAMILY OF RIBOSOMAL PROTEINS
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcements
CC or send an email to license@isb-sib.ch)
CC
DR EMBL: D17510; BAA04342.1; -
DR Mendei: 15014; PINH; rpl20.1.
DR InterPro: IP001081; RIBOSOMAL_L20.
DR Pfam: PF00454; RIBOSOMAL_L20; 1.
DR PRINTS: PK00062; RIBOSOMAL_L20.
DR Problem: P0002489; RIBOSOMAL_L20; 1.
DR P0051P; P00043; RIBOSOMAL_L20; 1.
KW Ribosomal protein; RNA-binding; Chloroplast.
FT INIT_MET 0 BY SIMILARITY.
SQ SEQUENCE 114 AA; 13745 MW; 38FF4221658E0E73 CRC64;

Query Match
Best Local Similarity 71.4%; Score 35; DB 1; Length 118;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 LRRMMVT 9
DB 56 LRRIMVT 62
|||||

RESULT 6
P152_METTH
ID P152_METTH STANDARD; PRT: 186 AA.

DR PRINTS: P00062; RIBOSOMAL_L20.
DR Problem: P0002486; RIBOSOMAL_L20; 1.
DR PROSITE: P00037; RNA-binding; Chloroplast.
KW Ribosomal protein; RNA-binding; Chloroplast.
FT INIT_MET 0 BY SIMILARITY.
SQ SEQUENCE 115 AA; 13475 MW; A91B680F1D777A39 CRC64;

Query Match
Best Local Similarity 71.4%; Score 35; DB 1; Length 186;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLPPMMVT 9
DB 81 DILPPMMVT 89
|||||

RESULT 7
MFGM_PIG
ID MFGM_PIG STANDARD; PRT: 409 AA.
AC P79385;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Lactadherin (Milk fat globule EGF factor 8) (MFG EB) (MFGM) (Sperm
DE surface protein SP47) (PB47).
GN MFGEB.

```


OS Sus scrofa (pig)
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 CC NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Enslin M.A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBT databases.
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA
 CC BINDING PROTEIN.
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
 CC SPERMATOZOAN
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C1-MAINS
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/autodoc/>
 CC or send an email to license@sib-sib.ch).

EMBL: Y11683; CAA72379.1;
 DR HSSP: P00740; 1EDM.
 DR InterPro: IPR000421; PA58_C
 DR Pfam: PF00008; EGF; 2.
 DR SMART: SM0181; EGF; 2
 DR SMART: SM00241; PA58C; 2
 DR PROSITE: PS00022; EGF_1; 2.
 DR PROSITE: PS01186; EGF_2; 2
 DR PROSITE: PS01285; PA58C_1; 2.
 DR PROSITE: PS01286; PA58C_2; 2.
 KW Glycoprotein; Repeat; EGF-like domain.
 FT DOMAIN 2 41 EGF-LIKE 1.
 FT DOMAIN 44 88 EGF-LIKE 2.
 FT DOMAIN 91 247 F5/8 TYPE C 1.
 FT DOMAIN 252 409 F5/8 TYPE C 2.
 FT SITE 67 69 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 6 17 BY SIMILARITY.
 FT DISULFID 11 29 BY SIMILARITY.
 FT DISULFID 31 40 BY SIMILARITY.
 FT DISULFID 91 247 BY SIMILARITY.
 FT DISULFID 234 248 BY SIMILARITY.
 FT DISULFID 252 409 BY SIMILARITY.
 FT CARBOHYD 41 41 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 372 372 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 409 AA; 45725 MW; 80C07AF9029927A CRC64;

Query Match 71.4%; Score 35; DB 1; Length 409;
 Best Local Similarity 88.9%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLLRRMWT 9
 DB 152 NLLRRMWT 160

RESULT 8
 ID OD01_RUCAI STANDARD; PPT; 909 AA.
 AC P57388.
 DT 16-OCT-2001 (rel 45, last sequence update)
 DT 16-OCT-2001 (rel 40, last sequence update)
 DT 16-OCT-2001 (rel 40, last annotation update)
 DE 2 oxoglutarate dehydrogenase E1 component (EC 1.2.4.2) (Alpha-
 DE ketoglutarate dehydrogenase).
 GN SUCA OR B0302.

OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 CC symbiotic bacterium).
 CC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 CC NCBI_TaxID=118099;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN TOKYO 1998;
 RX MEDLINE=20445173; PubMed=10991077;
 RA Shirogane S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 FT "Genomic sequence of the endosymbiotic bacterial symbiont of aphids
 CC Buchnera sp. APS.";
 RL Nature 407:81-86(2000).
 CC -1- FUNCTION: THE 2-OXOGLUTARATE DEHYDROGENASE COMPLEX CATALYZES THE
 CC OVERALL CONVERSION OF 2-OXOGLUTARATE TO SUCCINYL-CoA & CO(2). IT
 CC CONTAINS MULTIPLE COPIES OF 3 ENZYMIC COMPONENTS: 2 OXOGLUTARATE
 CC DEHYDROGENASE (E1), LIPOYL-LIP-AMINOTRANSFERASE (E2) AND
 CC LIPOAMIDE DEHYDROGENASE (E3) (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 2-oxoglutarate + lipoamide = S-
 CC succinylidihydrolipoamide + CO(2).
 CC -1- COFACTOR: THIAMINE PYROPHOSPHATE (BY SIMILARITY).
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/autodoc/>
 CC or send an email to license@sib-sib.ch).

EMBL: AP001118; BAB13011.1;
 DR InterPro: IPR01017; EL_dh.
 DR InterPro: IPR002986; PPTA.
 DR Pfam: PF05676; PL_dhdrog; 1.
 KW Glycolysis, Oxidoreductase, Flavoprotein; Thiamine pyrophosphate;
 KW Complete proteome.
 SQ SEQUENCE 935 AA; 105864 MW; A6739F8CD9F9CF8B CRC64;

Query Match 71.4%; Score 35; DB 1; Length 909;
 Best Local Similarity 55.6%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLLRRMWT 9
 DB 157 NLLRRMWT 165

RESULT 9
 ID RK20_CHLRE STANDARD; PPT; 111 AA.
 AC P26565;
 DT 01-AUG-1992 (rel 23, created)
 DT 01-DEC-1992 (rel 24, last sequence update)
 DT 16-OCT-2001 (rel 40, last annotation update)
 DE Chloroplast 50S ribosomal protein L20.
 GN RPL20.
 OS Chlamydomonas reinhardtii.
 CC Chloroplast.
 CC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 CC Chlamydomonadales; Chlamydomonadales;
 CC NCBI_TaxID=3055;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-2137;

QY 1 NLLRRMWT 9
 DB 157 NLLRRMWT 165

RESULT 8
 ID OD01_RUCAI STANDARD; PPT; 909 AA.
 AC P57388.
 DT 16-OCT-2001 (rel 45, last sequence update)
 DT 16-OCT-2001 (rel 40, last sequence update)
 DT 16-OCT-2001 (rel 40, last annotation update)
 DE 2 oxoglutarate dehydrogenase E1 component (EC 1.2.4.2) (Alpha-
 DE ketoglutarate dehydrogenase).
 GN SUCA OR B0302.

the European Bioinformatics Institute. There are no restrictions on its use by copyright institutions as long as its content is not modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: X62566; CAA43439.1; -
 PIR: S18026; P5FM20.
 Model: 4049; Chlre: rpl20; 1.
 InterPro: IPR001081; Ribosomal_L20.
 Pfam: PF00454; Ribosomal_L20; 1.
 PRINTS: PR00062; RIBOSOMAL_L20.
 PRODOM: P00023489; Ribosomal_L20; 1.
 PROSITE: PS00933; RIBOSOMAL_L20; 1.
 Ribosomal protein: rRNA-binding; Chloroplast.
 INT_MOT 0 BY SIMILARITY.
 SEQUENCE 111 AA: 13417 MW: 62034729015 FRP: 0.064.

Query Match 69.4%; Score 34; DB 1; Length 111;
 Best local Similarity 83.4%; Pred. No. 5;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0.

QY 4 RRMWT 9
 |||||
 DB 57 RRMWT 62

RESULT 10
 NP5C_D10GR STANDARD; PRT: 699 AA.
 AC Q32131;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DI 15-JUL-1998 (Rel. 36, Last annotation update)
 DE NADH-plastoquinone oxidoreductase chain 5, chloroplast (EC 1.6.5.3) (Fragment).
 DE NADH.
 GN NADH.
 OS Digitalis grandiflora (Yellow foxglove).
 CC Chloroplast.
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot,
 CC Asteridae; euasterids I; Lamiales; Veronicaceae; Digitalis.
 CC NCBI_TaxID: 38791;
 CC |||
 CC SEQUENCE FROM N.A.
 CC Olmstead R.G., Reeves P.A.
 CC "Evidence for the polyphyly of the Scrophulariaceae based on
 CC chloroplast rbcL and ndhF sequences."
 CC Ann. Mo. Bot. Gard. 82:174-193(1995).
 CC -!- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE -> NADP(+) + PLASTOQUINOL.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: U36494; MAA84203.1; -
 Model: 2414; D1Gqrindhp; 1.
 InterPro: IPR001750; oxidored_q1.
 InterPro: IPR002128; oxidored_q1_N.
 InterPro: IPR001516; oxid-red_q1_N.
 Pfam: PF00361; oxidored_q1; 1.
 Pfam: PF01010; oxidored_q1_C; 1.
 Pfam: PF00662; oxidored_q1_N; 1.
 oxidoreductase; NAD, Plastoquinone; Chloroplast.
 FT NON_TER 1
 FT NON_TER 699
 SEQUENCE 699 AA: 79309 MW: 5685FBA561C63D01 CRC64.

Query Match 69.4%; Score 34; DB 1; Length 199;
 Best local Similarity 85.7%; Pred. No. 34;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0.

QY 1 NLLRRMW 7
 |||||
 DB 26 NLLRRMW 32

RESULT 11
 YLB5_CAEEL STANDARD; PRT: 1281 AA.
 AC P46580;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DI 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 146.8 kDa protein C34E10.5 in chromosome III.
 GN C34E10.5
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Platyhelminthes; Platyhelminthes; Platyhelminthes; Platyhelminthes.
 CC NCBI_TaxID: 6239;
 CC |||
 CC SEQUENCE FROM N.A.
 CC STRAIN-BRISTOL N2;
 CC Kirsten J.;
 CC Submitted (JUN-1994) to the EMBL/GenBank/Tran databases.
 CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC -!- SIMILARITY: TO S-POMBE SKB1 AND YEAST YBK133C.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: U10402; AAA19067.1; -
 WormPe: C34E10.5; C001185.
 InterPro: IPR001841; ZnF_RING.
 Pfam: PF00097; ZF-C3HC4; 1.
 SMART: SM00184; RING; 1.
 PROSITE: PS00518; ZF_RING_1; 1.
 PROSITE: PS00589; ZF_RING_2; 1.
 Hypothetical protein; Transmembrane; Zinc-finger.
 FT TRANSMEM 69
 FT TRANSMEM 1165 1185 POTENTIAL.
 FT ZN_FING 1222 1267 RING-TYPE.
 SEQUENCE 1281 AA: 145773 MW: 7027275AAB540D04 CRC64;

Query Match 69.4%; Score 34; DB 1; Length 1281;
 Best local Similarity 55.8%; Pred. No. 65;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0.

QY 1 NLLPRMWT 9
 |||||
 DB 271 NLLPRMWT 279

RESULT 12
 ARGH_HUMAN STANDARD; PRT: 234 AA.
 AC P55145;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DI 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Arginine-rich protein.
 GN ARG.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia: Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. AND VARIANT CANCER APC-50
 EX MEDLINE=96211400; PubMed=8649864;
 FA Shridhar V., Rivard S., Shridhar R., Mullins C., Saki W.,
 RA Grignon D., Miller O.J., Smith D.I.;
 PT "A gene from human chromosome band 3p21.1 encodes a highly conserved
 PT arginine-rich protein and is mutated in renal cell carcinomas";
 RL Oncogene 12:1931-1939(1996);
 RN [2]
 RP VARIANT ARG-50.
 EX MEDLINE=97106922; PubMed=8971154;
 FA Shridhar V., Shridhar R., Rivard S., Shridhar R.M.,
 RA Pietraszkiewicz H., Pauley J., Pauley R., Grignon D., Saki W.,
 RA Miller O.J., Smith D.I.;
 PT "Mutations in the arginine-rich protein gene, in lung, breast, and
 PT prostate cancers, and in squamous cell carcinoma of the head and
 RT neck";
 RL Cancer Res. 56:5576-5578(1996);
 RN [3]
 RP VARIANTS IN PANCREATIC CANCERS.
 EX MEDLINE=97416783; PubMed=9174057;
 FA Shridhar V., Rivard S., Waka X., Shridhar R., Pauley R., Mullins C.,
 PA Beirnat J., Dugan M., Sarkar F., Miller O.J., Vaitkevicius V.K.,
 RA Smith D.I.;
 PT "Mutations in the arginine-rich protein gene (ARF) in pancreatic
 RT cancer";
 RL Oncogene 11:2213-2216(1997);
 CC -|- SIMILARITY: BELONGS TO THE APC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/anonymous/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M83751; AAB08753.1; ALT_INIT.
 DR MIM: 601916;
 DR Disease mutation.
 KW DOMAIN 38 55 POLY-APC.
 FT VARIANT 50 50 M -> R (IN CANCER).
 FT VARIANT 50 50 /FTID-VAR_010245.
 FT MISSING (IN CANCER).
 FT VARIANT 50 50 MISSING (IN CANCER).
 FT /FTID-VAR_010246.
 FT VARIANT 51 51 R -> K (IN CANCER).
 FT /FTID-VAR_010247.
 FT /FTID-VAR_010248.
 SQ SEQUENCE 234 AA; 26466 MW; 1048162F6967564 CRC64.

Query Match 67.34; Score 33; DB 1; Length 234;
 Best Local Similarity 71.48; Pred. No. 17;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 LRRMWT 9
 :|||:
 Db 53 MRPMTAT 50
 RESULT 13
 YJ95_CAEEL
 ID YJ95_CAEEL STANDARD; PRU; 468 AA.
 AC P49049;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Hypothetical 52 kDa protein T05E11.5 in Chromosome IV.
 GN T05E11.5
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Telodrilinae; Caenorhabditis.

OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 PC STRAIN PP1ST01.02;
 FA Kershaw J.;
 RA Kershaw J.;
 CC -|- SIMILARITY: TO YEAST YKL100C.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/anonymous/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: Z62751; CAA92751.1;
 DR WormPep; T05E11.5; CE06464;
 KW Hypothetical protein.
 SQ SEQUENCE 458 AA; 52793 MW; 00DFFACB679AA5F8 CRC64;

Query Match 67.49; Score 33; DB 1; Length 458;
 Best Local Similarity 55.68; Pred. No. 36;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 NITGMMWT 9
 :|||:
 Db 263 HLLRNHWIT 271
 RESULT 14
 Y4FD_RHISN
 ID Y4FD_RHISN STANDARD; PRU; 220 AA.
 AC P55442;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical 24.6 kDa protein Y4FD.
 GN Y4FD.
 OS Rhizobium sp. (strain NCP234).
 CC Plasmid sym NCP234a.
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97305956; PubMed 9163424;
 PA Freilich G.A., Fellay P., Raitore A., Proghron W., Vegerhoff A.,
 RA Perret X.;
 PT "Molecular basis of symbiosis between rhizobium and legumes";
 RL Nature 387:394-401(1997).
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -|- SIMILARITY: NONE OBSERVED.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/anonymous/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AF000072; AAB91661.1;
 DR Hypothetical protein; Transmembrane; plasmid.
 FT TRANSMEM 9 29 POTENTIAL.
 FT TRANSMEM 54 74 POTENTIAL.
 FT TRANSMEM 105 125 POTENTIAL.
 FT TRANSMEM 177 197 POTENTIAL.
 SQ SEQUENCE 220 AA; 24627 MW; 50411A4A84557381 CRC64;

Query Match 65.49; Score 32; DB 1; Length 220;

Best Local Similarity 57.1%; Prod. No. 25;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLLRRMW 7
Db :25 NVIRRLW 131

Search completed: September 5, 2002, 15:31:40
Job time: 475 sec

QY 3 LRRMWVT 9
Db 6 LRRMWVT 12

RESULTS 15

GB12_CHICK STANDARD: PRT: 354 AA.
AC P50147;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Guanine nucleotide-binding protein G(1), alpha-2 subunit (Adenylylate
DE cyclase-inhibiting G alpha protein).
GN GNA12.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae, Galliformes, Phasianidae, Phasianinae;
OC Gallus.
OX NCBI_TaxID-9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE 95121926; PubMed-7821804;
RA Kilbourne E.J., Galper J.B.;
RT "Cloning of cDNAs coding for the G alpha i1 and G alpha i2 G-proteins
RT from chick brain."
RL Gene 150:341-344(1994).
CC -!- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE
CC INVOLVED AS MODULATORS OR TRANSDUCERS IN VARIOUS TRANSMEMBRANE
CC SIGNALING SYSTEMS.
CC -!- FUNCTION: THE G(1) PROTEINS ARE INVOLVED IN HORMONAL REGULATION OF
CC ADENYLATE CYCLASE: THEY INHIBIT THE CYCLASE IN RESPONSE TO BETA-
CC ADRENERGIC STIMULI.
CC -!- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).
CC THE ALPHA CHAIN CONTAINS THE GUANINE NUCLEOTIDE BINDING SITE.
CC -!- SIMILARITY: BELONGS TO THE G-ALPHA FAMILY. SUBFAMILY 1
CC (G(1/G12)).
CC
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L24549; AAA65067.1; -.
DR BSSP: P10H24; IAS3.
DR InterPro: IPR001019; Gprotein_alpha.
DR InterPro: IPR001230; Prenyltn.
DR Pfam: PF00503; G-alpha.1.
DR PRINTS: PK00318; GPROTEIN.
DR SMART: SM00275; G-alpha.1.
KW GTP-binding; Transducer; ADP-ribosylation; Multiqene family;
KW Myristate; Palmitate; Lipoprotein.
FT INIT_MET 0 0 BY SIMILARITY.
FT LIPID 1 1 MYRISTATE (BY SIMILARITY).
FT LIPID 2 2 PALMITATE (BY SIMILARITY).
FT NP_BIND 39 46 GTP (BY SIMILARITY).
FT NP_BIND 200 204 GTP (BY SIMILARITY).
FT NP_BIND 269 272 GTP (BY SIMILARITY).
FT MOD_RES 178 178 ADP-RIBOSYL[1] (BY ACTION OF CTX).
FT MOD_RES 451 351 ADP-RIBOSYL[1] (BY ACTION OF IAP).
SQ SEQUENCE 354 AA; 40446 MW; D964549D9C5CC4F CRC64;

Query Match 65.3%; Score 32; DB 1; Length 354;
Best Local Similarity 57.1%; Prod. No. 42;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;



A:Reference number: S74211; MUID:97008954
 A:Accession: S74211
 A:Molecule type: mpna
 A:Residues: 1-427 <HVA>
 A:Cross-references: EMBL:X51845, NLE_g1532778, E128_CAA62537.1, F15-g1532779
 A:Accession: S78114
 A:Molecule type: protein
 A:Residues: 35, 85, 96, 110, 140, 165, 174, 179, 221, 242, 248, 277, 289, 295, 309, 337, 359, 410, 425, 427
 R:Kim, D.H.; Kanzo, C.; Mizokami, Y.
 Biochim. Biophys. Acta 1122, 293-311, 1992
 A:Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, from
 A:Reference number: S23926; MUID:92353107
 A:Accession: S24181
 A:Molecule type: protein
 A:Residues: 383-394 <KM>
 C:Superfamily: milk fat globule protein; discoidin I aminoterminal homology; PPF hemolys
 C:Keywords: blocked amino end; disulfide bond; glycoprotein; milk
 F1:18/omega1 signal sequence; #status predicted <SIG>
 F19:4-7/omega1 PAS 6/7 protein; #status experimental <MAT>
 F24:58/omega1: PPF homology <EG>
 F26:105/omega1: PPF homology <EG>
 F108:245/omega1: discoidin I amino terminal homology <N1>
 F200:427/omega1: discoidin I amino terminal homology <N2>
 F224:35,29-47,49-58,66-77,71,94,96,105/disulfide bonds; #status predicted
 F27:Binding site: carbohydrate (Ser) (covalent) #status experimental
 F34:Binding site: carbohydrate (Thr) (covalent) #status experimental
 F59:227/Binding site: carbohydrate (Asp) (covalent) #status experimental
 F109:245,252-256,276-278/disulfide bonds; #status experimental
 Query Match 99.8%; Score 44; DB 2; Length 427;
 Best Local Similarity 77.8%; Pred. No. 5.74;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NLLRRMWT 9
 I I I I I I I
 DB 170 NLLRRMWT 178
 RESULT 3
 S54218
 FlgA protein - Yersinia enterocolitica
 C:Species: Yersinia enterocolitica
 C:Date: 08-Jul-1995; #sequence_revision 21-Jul-1995; #text_change 29-Sep-1999
 C:Accession: S54218
 R:Faucquier, A.; Allaoui, A.; van Elsen, A.; Cornelis, G.; Boilen, A.
 submitted to the EMBL Data Library, February 1995
 A:Description: Clustering of flagellar genes around invA, the Yersinia enterocolitica in
 A:Reference number: S54213
 A:Accession: S54218
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-224 <TAG>
 A:Cross-references: EMBL:Z48169; NDB:q793891; PDB:CAA88191.1; PDB:q793897
 C:Superfamily: flagellar basal body p-ring protein flgA
 Query Match 77.6%; Score 38; DB 2; Length 224;
 Best Local Similarity 62.5%; Pred. No. 5.7;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NLLRRMWT 8
 I I I I I I I
 DB 159 NLLRRMWT 166
 RESULT 4
 AC0219
 flagella basal body P ring formation protein FlgA [imported] - Yersinia pestis (strain G
 C:Species: Yersinia pestis
 C:Date: 02-Nov-2001; #sequence_revision 02-Nov-2001; #text_change 09-Nov-2001
 C:Accession: AC0219
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tittball, K.W.; Holt-Bu, M.T.G.; Protheroe, M.R.

deno-Farrado, A.M.; Chillingworth, T.; Green, A.; Davies, P.M.; Davis, P.; Dodson, G.
 il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrcl
 Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0531, MUID:21470413; PMID:11586360
 A:Accession: AC0219
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-232 <KOR>
 A:Cross-references: GB:AL590842; FIDR:CAAC90615.1; FID:q15979821; GENE:GN00175
 C:Genetics:
 A:Gene: flgA
 C:Superfamily: flagellar basal body P-ring protein flgA

Query Match 79.5%; Score 36; DB 2; Length 202;
 Best Local Similarity 62.5%; Pred. No. 15;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NLLRRMW 8
 I I I I I I I
 DB 167 NLLRRMW 174
 RESULT 5
 T51197
 hypothetical protein B7M.40 [imported] Neurospora crassa
 C:Species: Neurospora crassa
 C:Date: 29-Jul-2000; #sequence_revision 28-Jul-2000; #text_change 28-Jul-2000
 C:Accession: T51197
 R:Schulte, B.; Aign, V.; Becheisel, J.; Brandt, P.; Fartmann, B.; Holland, K.; Nyakatu
 submitted to the Protein Sequence Database, July 2000
 A:Reference number: Z25286
 A:Accession: T51197
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-390 <SCH>
 A:Cross-references: EMBL:AF300218; GDB:683016; NDB:p.B7M.40
 A:Experimental source: BAC clone B7M4; strain OR74A
 C:Genetics:
 A:Gene: NCSP:B7M4.40
 A:Map position: 6
 A:Introns: 56/3
 C:Superfamily: Neurospora crassa hypothetical protein B7M.40

Query Match 73.5%; Score 36; DB 2; Length 390;
 Best Local Similarity 85.7%; Pred. No. 24;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 NLLRRMW 7
 I I I I I I I
 DB 272 NLLRRMW 278

RESULT 6
 A13148
 conserved hypothetical protein bae3 [imported] - Agrobacterium tumefaciens (strain C5
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002; #sequence_revision 11-Jan-2002; #text_change 11-Jan-2002
 C:Accession: A13148
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
 erage, G.; Gillet, W.; Grant, G.; Guenther, D.; Kuryavin, T.; Levy, R.; Li, M.; Mehl
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Hiddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon Kam
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58
 A:Reference number: AB2577; PM:011743193
 A:Accession: A13148
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-405 <KUP>


```

Query Match      71.4%  Score 39; DB 2; Length 19;
Best local Similarity 71.4%  Prod. No. 12;
Matches 5; Conservative 2; Mismatches 0;
Ov 3 LPRMMWT 9; Caps 0;

```

```

|||||
Db 57 LRRRWIT 63

RESULT 11
G69069
hypothetical protein MTH152 - Methanobacterium thermoautotrophicum (strain Delta H)
C:Species: Methanobacterium thermoautotrophicum
C>Date: 10-Sep-1999 #sequence_revision 16-Sep-1999 #text_change 21-Jul-2000
C:Accession: G69069
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
Qiu, D.; Spadafora, R.; Vicarel, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.;
El, S.; Church, G.M.; Daniels, C.J.; Mac, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum delta H - funct
A:Reference number: A69000; MUID:98047514
A:Accession: G69069
A:Status: preliminary
A:Suprafamily: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1186 <MTH>
A:Cross-references: GR:AF000803; GR:AF000805; NID:32521179; FIDR:AAB4558.1; FID:3252119
C:Genetics:
A:Gene: MTH152
A:Start codon: GTG
C:Superfamily: Bacillus subtilis hypothetical protein ywrf

Query Match 71.4%; Score 35; DB 1; Length 186;
Best Local Similarity 55.6%; Pred. No. 18;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NLLRRMWIT 9
|||||
Db 81 DIFRRMWIT 89

RESULT 12
T11743
pP47 protein - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T11743
R:Esslin, M.; Vogel, T.; Calvete, J.J.; Thole, H.H.; Schmidtke, J.; Matsuda, T.; Toepfe
biol. Reprod. 58, 1057-1064, 1998
A:Title: Molecular cloning and characterization of P47, a novel boar sperm-associated 22
A:Reference number: 217325; MUID:98206817
A:Accession: T11743
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-409 <ENS>
A:Cross-references: EMBL:Y11683; NID:3252227; FIDR:CAA2379.1; FID:3252228
A:Experimental source: testis
C:Function:
C:Description: may be involved in membrane remodeling and/or function as a zona pellucida
C:Superfamily: unassigned EGF-related proteins; EGF homology
F:6-40/domain: EGF homology <EGF>

Query Match 71.4%; Score 25; DB 2; Length 499;
Best Local Similarity 88.9%; Pred. No. 49;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLLRRMWIT 9
|||||
Db 152 NLLRRMWIT 160

RESULT 13
G84965
oxoglutarate dehydrogenase (lipoamide) (EC 1.2.4.2) cl component [imported] - Buchnera sp.
C:Species: Buchnera sp.
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Genetics:

```

```

C:Accession: G84965
P:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A:Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp.
A:Reference number: A84930; MUID:20445173
A:Accession: G84965
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1999 <STO>
A:Cross-references: GR:AP000308; GSPPR:GN00144
A:Experimental source: strain APS
C:Genetics:
A:Gene: sucA; B0492
C:Superfamily: oxoglutarate dehydrogenase (lipoamide); thiamin pyrophosphate-binding
C:Keywords: oxoglutarate
Query Match 71.4%; Score 35; DB 2; Length 909;
Best Local Similarity 55.6%; Pred. No. 87;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 NLLRRMWIT 9
|||||
Db 157 NLLRRMWIT 165

RESULT 14
R5RM20
ribosomal protein L20, chloroplast - Chlamydomonas reinhardtii chloroplast
C:Species: Chlamydomonas reinhardtii
C>Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 22-Jun-1999
C:Accession: S18026
R:Yu, W.; Zhang, D.; Spreitzer, R.J.
submitted to the EMBL data library, October 1991
A:Description: Sequences of the trnS and rpl20 genes of the Chlamydomonas reinhardtii
A:Reference number: S18026
A:Accession: S18026
A:Molecule type: DNA
A:Residues: 1112 <YCW>
A:Cross-references: EMBL:X62566; NID:311463; FIDR:CAA4439.1; FID:411464
A:Experimental source: strain 2137 mt+
C:Genetics:
A:Gene: rpl20
A:Superfamily: Escherichia coli ribosomal protein L20
C:Keywords: chloroplast, protein biosynthesis, ribosome

Query Match 69.4%; Score 34; DB 1; Length 112;
Best Local Similarity 83.3%; Pred. No. 17;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 4 PRMWIT 9
|||||
Db 58 PRMWIT 63

RESULT 15
T24743
hypothetical protein T99E11.10 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jan-2000
C:Accession: T24743
R:McLay, K.
submitted to the EMBL data library, October 1996
A:Reference number: Z19930
A:Accession: T24743
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1338 <WIL>
A:Cross-references: EMBL:Z81147; FIDR:CAB03547.1; GSFDL:GN00019; CESP:T09E11.10
A:Experimental source: clone T09E11
C:Genetics:

```

A:Gene: CESP:T09E11.10
 A:Map position: 14 154 155 156 157 158 159 160 161 162 163
 A:Title: 157: 14 154 155 156 157 158 159 160 161 162 163
 C:Superfamily: Caenorhabditis elegans hypothetical protein T09F5.1

Query Match 69.4% Score 34 DB 2 Length 338
 Best Local Similarity 55.6% Pred No. 51
 Matches 5 Conservative 3 Mismatches 1 Indels 0 Gaps 0

QY 1 NLLRRMWT 9
 1:11: 1:1
 Db 96 NVLRKTWMT 104

Search completed: September 5, 2002, 15:28:50
 Job time: 355 sec



GenCore version 4.5
Copyright (c) 1994 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002 16:31:09, search time: 122.40 seconds
(without alignments)
12,673 Million cell updates/sec

Title: US-09-744-804-36

Perfect score: 40

Sequence: 1 NLLRRMMWT 9

Scoring table: P100NM02

Gapop 10.0, Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_orqanelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_protist:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriophage:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	100.0	335	4 Q9BT19	Q9bt19 homo sapien
2	41	83.7	1928	4 G96736	G96736 homo sapien
3	41	83.7	2392	4 Q95001	Q95001 homo sapien
4	41	83.7	2412	4 Q92516	Q92516 homo sapien
5	36	73.5	340	3 Q9P3C0	Q9p3c0 neurospora
6	36	73.5	3071	10 Q9SND0	Q9snd0 arabidopsis
7	35	71.4	1292	10 Q91G50	Q91g50 cryza sativ
8	34	69.4	232	5 Q95613	Q95613 plasmodium
9	34	69.4	329	4 Q96158	Q96158 homo sapien
10	34	69.4	338	5 Q02310	Q02310 cucurbitabdi
11	34	69.4	404	2 Q93C08	Q93c08 shigella bo
12	34	69.4	645	8 G46345	G46345 heliotropiu
13	34	69.4	660	16 Q97RR4	Q97rr4 streptococ
14	34	69.4	700	8 Q9TIT9	Q9tit9 tourneforti
15	34	69.4	706	8 Q9T148	Q9t148 rhamnus dav
16	34	69.4	707	5 Q23034	Q23034 caenorhabdi

ALIGNMENTS

RESULT 1

ID Q9BT19 PRELIMINARY: PRT: 335 AA.
AC Q9BT19;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DI 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DF 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
IP SMITHAP TO MILK FAT GLOBULIN-EGF FACTOR R PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
GX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RA Strausberg R.;
FI Submitted (JUN 2001) to the EMBL/GenBank/DDBJ databases.
CC -L- SIMILARITY: CONTAINS 1 EGF/8 TYPE C DOMAIN.
DR EMBL: BC003610; AAH03610.1; 1;
DR HSP: P08709; IHP9;
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001438; EGF_11.
DR InterPro: IPR000421; FA58.C.
DR PRINTS: PR00010; EGFBL00D.
DR SMART: SM00181; EGF; 1.
DR SMART: SM00001; EGF-like; 1.
DR SMART: SM00231; FA58.C; 2.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01285; FA58.C; 1.
DR PROSITE: PS01286; FA58.C; 2.
KW EGF-like domain; Glycoprotein. 20E94FEIM78DE25 CRC64;
SQ SEQUENCE: 335 AA; 37523 MW; 20E94FEIM78DE25 CRC64;

Query Match: 100.0%; Score 49; DB 4; Length 335;

Best Local Similarity: 100.0%; Pred. No. 0.11;

Matches: 9; Conservative: 0; Mismatches: 0; Indels: 0; Gaps: 0;

```

QY 1 NLLRRMWT 9
   111111111
Db 131 NLLRRMWT 139

RESULT 2
Q99736 Q99736 PRELIMINARY; PRT: 1928 AA.
AC Q99736;
DT 01-MAY-1999 (TrEMBLrel. 03, Created)
DI 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HSCGN1 (FRAGMENT).
GN HSCGN1.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
CX NCBI_TaxID:9606;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE: SKELETAL MUSCLE;
RX MEDLINE:97376024; PubMed 9244705;
RA Marion M.J., Varghese J., Aldana C.R., Qiu H., Chakraborty K.,
RA Hinebusch A.G.;
RT "Evidence that GCN1 and GCN20, translational regulators of GCN4,
RT function on elongating ribosomes in activation of eIF2alpha kinase
RT GCN2";
RL Mol. Cell. Biol. 17:4474-4489(1997)
DR EMBL: U77700; AAC51648.1; -.
DR InterPro: IPR000357; HEAT_repeat.
DR InterPro: IPR000847; HTH_LYSR.
DR PTam: PF02985; HEAT: 7
DR PROSITE: PS00077; HEAT_REPEAT: 3.
DR PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
FT NON_TER 1 1
SQ SEQUENCE 1928 AA; 21150 MW; 8B82070387A8F89896964;

Query Match 83.7%; Score 41; DB 4; Length 1928;
Best Local Similarity 87.5%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLLRRMWT 8
   11111111
Db 294 NLLRRMWT 291

RESULT 3
Q95001 Q95001 PRELIMINARY; PRT: 2302 AA.
AC Q95001;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DI 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE WUSC_HL_267011.3 PROTEIN (FRAGMENT).
GN WUSC_HL_267011.3.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
CX NCBI_TaxID:9606;
RN 11
RP SEQUENCE FROM N.A.
RC Kolling T., Clarke K., Bauer C., Morris M.;
RA "The sequence of Homo sapiens PAC clone 267011.3";
RL submitted (JUN-1998) to the EMBL/GenBank/FDB databases
[2]
GN Watson R.;
RA Submitted (JUN 1998) to the EMBL/GenBank/FDB databases.
RL EMBL: AC004812; AAC83183.1; -.
DR InterPro: IPR000357; HEAT_repeat.
DR InterPro: IPR000847; HTH_LYSR.
DR InterPro: IPR000663; Natri_peptide.

QY 1 NLLRRMWT 8
   11111111
Db 866 NLLRRMWT 873

Query Match 83.7%; Score 41; DB 4; Length 2412;
Best Local Similarity 87.5%; Pred. No. 27;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLLRRMWT 8
   11111111
Db 846 NLLRRMWT 853

Query Match 83.7%; Score 41; DB 4; Length 2492;
Best Local Similarity 87.5%; Pred. No. 27;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLLRRMWT 8
   11111111
Db 846 NLLRRMWT 853

RESULT 4
Q92616 Q92616 PRELIMINARY; PRT: 2412 AA.
AC Q92616;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DI 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA0219 PROTEIN (FRAGMENT).
GN KIAA0219.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
CX NCBI_TaxID:9606;
RN 11
RP SEQUENCE FROM N.A.
RC T-SSUE: BONE MARROW;
RX MEDLINE:97191544; PubMed 9039502;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayashi Y., Ohara O.,
RA Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT the coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
PL DNA Res. 3:321-329(1996).
DR EMBL: D86973; BAA13209.1; -.
DR InterPro: IPR000357; HEAT_repeat.
DR InterPro: IPR000847; HTH_LYSR.
DR InterPro: IPR000663; Natri_peptide.
DR PTam: PF02985; HEAT: 7.
DR PRINTS: PR00710; NATPEPTIDES.
DR PROSITE: PS00077; HEAT_REPEAT: 3.
DR PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
FT NON_TER 1 1
SQ SEQUENCE 2412 AA; 264339 MW; 64F0D6DD8EEF090R CRC64;

Query Match 83.7%; Score 41; DB 4; Length 2412;
Best Local Similarity 87.5%; Pred. No. 27;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLLRRMWT 8
   11111111
Db 866 NLLRRMWT 873

Query Match 83.7%; Score 41; DB 4; Length 2412;
Best Local Similarity 87.5%; Pred. No. 27;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLLRRMWT 8
   11111111
Db 866 NLLRRMWT 873

RESULT 5
Q9P3C0 Q9P3C0 PRELIMINARY; PRT: 490 AA.
AC Q9P3C0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DI 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CONSERVED HYDROPHILIC PROTEIN.
GN B7N4.40.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.

```

OX NCBI_TaxID=5141;
 RN [1]
 PP SEQUENCE FROM N.A.
 RA Schulte U., Aign V., Holtheisel J., Brandt P., Farman R., Holland R.,
 RA Nyakatura G., Mewes H.W., Mannhaupt G.
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DP EMBL: AL392218; CAB92235.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 390 AA; 44439 MW; E1BE513D57672BD9 CRC64;

Query Match 73.5%; Score 36; DB 3; Length 300;
 Best Local Similarity 85.7%; Pred. No. 42;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 NLLPRMW 7
 [111111]
 DB 272 NLLPRMW 278

RESULT 6
 Q9SND0 PRELIMINARY; PRT; 3071 AA.
 AC Q9SND0;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE HYPOTHETICAL 343.7 KDA PROTEIN.
 GN FLIC1.220.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Barques M., Collado M.C., Navarro P., Toral J., Perez-Alonso M.,
 RA Mewes H.W., Fodor S., Lemcke K.E.N., Quetier F.,
 RA Salanoubat M.
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL 3401111 (AF345363) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL132976; CAB62317.1; -;
 DR InterPro: IPR000005; HTHARC.
 DP PPOSITE: P500041; HTH_APAC_FAM11_Y_1; ENRNWML1.
 KW Hypothetical protein.
 SQ SEQUENCE 3071 AA; 343688 MW; 4A3F7C9F71C92A25 CRC64;

Query Match 73.5%; Score 36; DB 10; Length 3071;
 Best Local Similarity 75.0%; Pred. No. 3,26+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY 1 NLLPRMW 8
 [111111]
 DB 601 NLLPRMW 608

RESULT 7
 Q9LG50 PRELIMINARY; PRT; 1292 AA.
 AC Q9LG50;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE ESTS AU078742(11888).
 OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. NIPPONBARE;
 RA Susaki T., Matsumoto T., Yamamoto K.;
 RL "Oryza sativa nipponbare (JA3) genomic DNA, chromosome 1, PAC
 clone: P0599B11";
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP002817; BAB03441.1; -;
 DR InterPro: IPR00767; Disease_resist.
 DR InterPro: IPR001611; LRR.
 DR InterPro: IPR004592; LRR_out.
 DE 20xRPT5; IPR002182; NB-ARC.
 DR Pfam: PF00560; LRR; 14.
 DR Pfam: PF00931; NB-ARC; 1.
 DR PRINTS: PR00364; DISEASERISIT.
 DR SMART: SM00370; LRR; 5.
 SQ SEQUENCE 1292 AA; 144595 MW; 3DBAR30DE947E767 CRC64;

Query Match 71.4%; Score 35; DB 10; Length 1292;
 Best Local Similarity 62.5%; Pred. No. 2,26+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Caps 0;

QY 1 NLLRRMW 8
 [111111]
 DB 1158 NLLRRMW 1165

RESULT 8
 Q25913 PRELIMINARY; PRT; 232 AA.
 AC Q25913;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE MALAKIA ANTIGEN (FRAGMENT).
 GN 226 32 12.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MALAYAN CAMP;
 RA MELIHL 2411952, Pubmed 1723149;
 RA Cheng Q., Jones G., Liu E.X., Kidson C., Saul A.;
 RL "Identification of a common plasmodium falciparum epitope (rPE)
 RT recognized by a pan-specific inhibitory monoclonal antibody";
 RL Mol. Biochem. Parasitol. 49:73-82(1991).
 DE EMBL: S53277; AAA29542.1; -;
 PE NON_TER
 SQ SEQUENCE 232 AA; 26506 MW; 7504991064697Ag CRC64;

Query Match 69.4%; Score 34; DB 5; Length 232;
 Best Local Similarity 71.4%; Pred. No. 62;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Caps 0;

QY 1 NLLRRMW 7
 [111111]
 DB 124 NLLRRMW 130

RESULT 9
 Q96L58 PRELIMINARY; PRT; 329 AA.
 AC Q96L58;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

```

DE BETA-1,3-GALACTOSYLTRANSFERASE-6.
GN B3GALT6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi.
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID 9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rafi X., Zhou B., Brown J.K., Hennet T., Esko J.D.,
RT "Biosynthesis of the linkage region of glycosaminoglycans: Cloning and
PT activity of galactosyltransferase II, the sixth member of the beta-
RT 1,3-galactosyltransferase family (beta3GALT6).";
RL Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AY050570; AAL11442.1;
KW Transmembrane Glycosyltransferase.
SQ SEQUENCE 329 AA; 37123 MW; 3753461B00DEA113 CRC64;

Query Match 69.4%; Score 74; DB 4; Length 329;
Best Local Similarity 85.7%; Pred. No. 87;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLLRRMW 7
   | | | | |
DB 2 NLLRRW 8

RESULT 10
ID 002310 PRELIMINARY; PRT; 338 AA.
AC 002310;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE TOPEL110 PROTEIN.
GN TOPEL110.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nemata; Chordata; Crustacea; Phlebotomidae;
OC Rhabdida; Polidoridae; Caenorhabditis
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Melay K.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA none;
RX MEDLINE 95069613; PubMed-9851916;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: Z81147; ZAB04537.1;
DR InterPro: IP002659; Galactosyl_T.
DR Pfam: PF01762; Galactosyl_T; 1.
SQ SEQUENCE 338 AA; 38916 MW; 247562F4024162B CRC64.

Query Match 69.4%; Score 34; DB 5; Length 338;
Best Local Similarity 55.6%; Pred. No. 90;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLLRRMW 9
   | | | | |
DB 96 NLLRRTW 104

RESULT 11
Q93CUB
ID Q93CUB PRELIMINARY; PRT; 409 AA.
AC Q93CUB;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

```

```

DE O-ANTIGEN FLIPPASE.
GN WZX.
OS Shigella boydii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_TaxID=621;
RN [1]
RP SEQUENCE FROM N.A.
RA MEHLNF21481970; PubMed-11598067;
RA Wang L., Qu W., Reeves P.K.;
RT "Sequence Analysis of Four Shigella loci: Implication
RT for Escherichia coli and Shigella Relationships.";
RL Infect Immun 69:6924-6930(2001).
DR EMBL: AF402312; AAL27316.1;
SQ SEQUENCE 409 AA; 47877 MW; 875B920EAC5CB1A CRC64;

Query Match 69.4%; Score 34; DB 2; Length 409;
Best Local Similarity 62.5%; Pred. No. 116-02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLLRRMW 8
   | | | | |
DB 198 NLLRRW 205

RESULT 12
O46945
ID 046945 PRELIMINARY; PRT; 645 AA.
AC 046945;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT F (FRAGMENT).
GN NdhF.
OS Heliotropium arborescens.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots;
OC Asteridae; euasterids I; euasterids I; incertae sedis; Boraginaceae;
OC Heliotropium.
OX NCBI_TaxID=28508;
RN [1]
RP SEQUENCE FROM N.A.
RA Ferguson D.M.;
RT "Phylogenetic analysis and circumscription of Hydrophyllaceae based on
RT ndhF sequence data.";
RL Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Schranz E., Olmstead R.G.;
KL Submitted (JUL 1997) to the EMBL/GenBank/DDBJ databases.
CC -1- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE -> NAD(+) + PLASTOQUINOL.
DR EMBL: AF014000; AAB94517.2;
DE InterPro: IPR001750; Oxidored_q1.
DR InterPro: IPR001218; Oxidored_q1_C.
DR InterPro: IPR001516; Oxidored_q1_N.
DE Pfam: PF00361; Oxidored_q1; 1.
DE Pfam: PF01810; Oxidored_q1_C; 1.
DE Pfam: PF06462; Oxidored_q1_N; 1.
KW Chloroplast; NAD; Oxidoreductase; Plastocyanine.
FT NON_TER 1
FT NON_TER 645
SQ SEQUENCE 645 AA; 72688 MW; 1CD1415H78EAD194 CRC64;

Query Match 69.4%; Score 34; DB 8; Length 645;
Best Local Similarity 100.0%; Pred. No. 1.76-02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LRRMW 8
   | | | | |
DB 28 LRRMW 33

```


RT "Phylogenetic analysis and circumscription of Hydrophyllaceae based on
ndhF sequence data.";

RL Syst. Bot. 0:0-0(1998).
CC 1 CATALYTIC ACTIVITY: NADH + PLASTOQUINONE NAD(+) + PLASTOQUINOL.
DR EMBL: AF047813; AAF22420.1; -;
DR InterPro: IPR001750; oxidored_q1.
DR InterPro: IPR00128; oxidored_q1_C.
DR InterPro: IPR001516; oxidored_q1_N.
DR Pfam: PF00361; oxidored_q1; 1.
DR Pfam: PF01010; oxidored_q1_C; 1.
DR Pfam: PF00662; oxidored_q1_N; 1.
KW Chloroplast; NAD; Oxidoreductase; plastoquinone.
FT NON_TER 1 700
SQ SEQUENCE 700 AA: 96BA728DAA7570C CRC64;

Query Match 69.4%; Score 24; DB 9; Length 700;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q7 LRRMWV 8
|||||
DE 28 LRRMWV 33

RESULT 15

Q9TL48 PRELIMINARY: PRT: 706 AA.

ID Q9TL48
AC Q9TL48;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NADH DEHYDROGENASE SUBUNIT F (FRAGMENT).
GN NDH.
OS Rhizobium davarica.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosales;
OC Ericales 1; Rosales; Rhamnaceae; Rhamnus.
OX NCBI_TaxID:105902;
RN [1]
RF SEQUENCE FROM N.A.
FX MEDLINE 20343987; PubMed-10877943;
FA Olmstead R.G., Kim K.J., Jansen R.K., Wagstaff S.J.;
RT "The phylogeny of the asteridae sensu lato based on chloroplast ndhF
gene sequences.";

RL Mol. Phylogenes. Evol. 16:96-112(2000).
CC 1 CATALYTIC ACTIVITY: NADH + PLASTOQUINONE NAD(+) + PLASTOQUINOL.
DR EMBL: AF130225; AAF08181.1; -;
DR InterPro: IPR001750; oxidored_q1.
DR InterPro: IPR00128; oxidored_q1_C.
DR InterPro: IPR001516; oxidored_q1_N.
DR Pfam: PF00361; oxidored_q1; 1.
DR Pfam: PF01010; oxidored_q1_C; 1.
DR Pfam: PF00662; oxidored_q1_N; 1.
KW Chloroplast; NAD; Oxidoreductase; plastoquinone.
FT NON_TER 1 706
SQ SEQUENCE 706 AA: 896A/EA5448830b CRC64;

Query Match 69.4%; Score 24; DB 9; Length 706;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q7 LRRMWV 8
|||||
DE 28 LRRMWV 33

Search completed: September 5, 2002, 15:31:01

RESULT 13

Q97RR4 PRELIMINARY: PRT: 660 AA.

ID Q97RR4
AC Q97RR4;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CATION-TRANSPORTING ATPASE, EL-E2 FAMILY.
GN SP0724.

OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.

OX NCBI_TaxID=1313;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-TIGR4;

FX MEM INF-21357200; PubMed-111463014;

RA Tettelin H., Nelson K.F., Paulson I.T., Pison J.A., Read T.D.,
RA Tettelin H., Nelson K.F., Paulson I.T., Haft D.H., Deason R.J.,
RA Peterson S., Heidelberg J., Bejery R.T., Haft D.H., Deason R.J.,
RA Durkin A.S., Cavin M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Dwayan L.A., White O., Salzberg S.L., Lewis M.R., Kadane D.,
RA Holzapfel F., Khouri H., Wolf A.M., Tarrback T.P., Hanson C.L.,
RA McDonald L.A., Feldblyum I.V., Anguillo S., Dickinson T., Hickey E.K.,
RA Holt J.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
pneumoniae.";

RL Science 293:498-506(2001).

DR EMBL: AF007381; AAK74870.1; -

DR TIGR: SP0729; -

DR InterPro: IPR001756; Cu-ATPase.

DR InterPro: IPR001757; EL-E2 ATPase.

DR InterPro: IPR001454; Hydrolase.

DR InterPro: IPR00150; Hypothet_cof

DR Pfam: PF00122; EL-E2 ATPase; 1.

DR Pfam: PF00702; Hydrolase; 1.

DR PRINTS: PR00119; CUAATPASE

DR PRINTS: PR00943; CUAATPASE

DR PROSITE: PS00154; ATPASE_EL_E2; UNKNOWN_1.

DR PROSITE: PS01229; COF_E2; UNKNOWN_1.

KW Complete proteome.

SQ SEQUENCE 660 AA: 70670 MW: 0A9826A1A6479AC CRC64;

Query Match 69.4%; Score 34; DB 16; Length 660;

Best Local Similarity 55.6%; Pred No. 1.7e+02;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLLRRMWV 9

|||||

DB 19 SLLQREWIT 27

RESULT 14

ID Q9TIT9

AC Q9TIT9 PRELIMINARY: PRT: 790 AA.

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE NADH DEHYDROGENASE (FRAGMENT)

GN NDH.

OS Tournefortia acutiflora.

OC Chloroplast.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Astoridaceae; Gerardiaceae; Hydrophyllaceae; Tournefortia.

OX NCBI_TaxID=79404;

RN [1]

RP SEQUENCE FROM N.A.

RA Ferguson D.M.;

Job time: 466 sec

XX DB WPI: 2000-205463/18.
 XX Tumor associated antigen peptides, especially derived from uroplakin,
 PT useful as vaccines to prevent or cure cancers including breast,
 PT bladder, prostate, pancreas, ovary, thyroid, colon and stomach -
 XX
 XX Claim 17: Page 99; 113pp; English.
 XX Tumor associated antigen peptides (TAA) may be used for the
 CC treatment, prevention and cure of cancer or cancer metastases. The
 CC cancer may be breast, bladder, prostate, pancreas, ovary, thyroid,
 CC colon, stomach, head or neck cancer or a carcinoma. The tumor
 CC associated antigens are presentable to the immune system by HLA A2
 CC molecules and are generally between 8 to 10 amino acids in length.
 CC The amino acids located at positions 2 and 9 of the tumor associated
 CC antigens are the anchor residues which participate in the binding to
 CC MHC class I molecules, more specifically HLA-A2. More tumor
 CC associated antigens are described in GenBank records AAY82806-Y82882.
 CC Those tumor associated antigens described in records AAY82806-Y82824
 CC and AAY82855-Y82869 are derived from uroplakin, such as Uroplakin II,
 CC uroplakin Ia, Uroplakin III and Uroplakin Ib. Those described in
 CC records AAY82825-Y82829 are derived from prostate specific antigen
 CC (PSA). Those described in records AAY82830-Y82835 are derived from
 CC prostate specific membrane antigen (PSMA). Those described in
 CC records Y82836-AAY82839 are derived from prostate acid phosphatase
 CC (PAP). Those described in records AAY82840-Y82846 are derived
 CC from Lactadherin (BA-46). Those described in records AAY82847-Y82854
 CC are derived from Mucin and those described in records AAY82871-Y82882
 CC are derived from Teratocarcinoma derived growth factor (TRIPG-1).
 XX
 XX Sequence 9 AA;
 SQ

Query Match 100.0%; Score 49; DB 21; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLLRRMMWT 9
 Db 1 nllrrmmwt 9

RESULT 2
 AAR77252
 ID AAR77252 standard; Protein: 387 AA

XX AAR77252;

XX 21-NOV-1995 (first entry)

XX BMFG 46 kDa antigen.

XX BMFG; human milk fat globule; antigen; immunogen; vaccine; virucide;
 KW epithelium; tumor; breast cancer; monoclonal antibody; MAb.

XX Homo sapiens.

XX W09515171-A.

XX 08-JUN-1995.

XX 05-DEC-1994; 94WO-US13967.

XX 03-DEC-1993; 93US-0162402.

XX (CANC-) CANCER RES FUND CONTRA COSTA.

XX Ceriani RL, Larocca DJ, Peterson JA;

XX WPI: 1995-215151/28.

XX N-PSDB, AAQ61198.

XX

PT 46 kD apparent molecular weight human milk fat globule antigen -
 PT used in assays to determine the presence of a cancerous tumour of
 PT epithelial origin, and in a vaccine against neoplastic tumours
 XX
 XX Claim 6: Page 46-47; 68pp; English.
 XX A complete cDNA sequence for the 46 kDa BMFG antigen, a major
 CC component of the apical surface of the normal breast epithelial
 CC cell, was obtained by PCR and RACE methods. cDNA clones can be
 CC used to prepare MABs for use in immunotherapy, immunohistochemistry,
 CC prognosis, diagnosis, imaging and therapy. Recombinant antigen can
 CC be expressed in prokaryotic or (glycosylated) in eukaryotic cells.
 XX
 XX Sequence 387 AA;
 SQ

Query Match 100.0%; Score 49; DB 16; Length 387;

Best Local Similarity 100.0%; Pred. No. 0.42;
 Matches 0; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NLLRRMMWT 9
 Db 131 nllrrmmwt 139

RESULT 3

AAY94453
 ID AAY94453 standard; Protein: 387 AA

XX AAY94453;

XX 11-SRP-2000 (first entry)

XX Human lactadherin protein.

XX Human; lactadherin; MCP-PR; anti-tumour; immune response;

KW exosome; dendritic cell.

XX Homo sapiens.

XX Key Location/Qualifiers

PT Peptide 1..23

PT Protein /label Secretion_signal

PT Binding-site 46..48

PT /label- Lactadherin

PT /label- Integrin_binding_site

XX EP1004664-A1.

XX 31-MAY-2000.

XX 24-NOV-1998; 98EP-0402025.

XX 24-NOV-1998; 98EP-0402025

XX (INRM) INRM INST NAT SANTE & RECH MEDICALE.

XX (CURI) INST CURIE.

XX WPI: 2000-352597/31.

XX N-PSDB; AAA27140.

XX Chimeric isolated (human) lactadherin polypeptide that functions as an

PT adaptor of cross-priming to eliminate pathogenic antigens, e.g. in

PT tumors -

XX Example 3; Page 12; 20pp; English.

XX Lactadherin protein was found in exosomes produced by dendritic cells.
 CC The protein is involved in the phagocytosis of particulate antigens by
 CC dendritic cells. Exosomes produced by dendritic cells exposed to
 CC tumour antigens induce potent immune responses. Lactadherin or variants
 CC of it may be used in the mediation of an immune response. Variants of

CC lactadherin may be used for inhibition and/or stimulation of the
 CC cross-priming of antigens and stimulation of the phagocytosis of
 CC antigens by dendritic cells. Compositions derived from lactadherin can
 CC also be used to monitor an immune response, more specifically a CTL
 CC (cytotoxic T-lymphocyte) response and also to produce CTLs specific
 CC for a selected antigen. The present sequence is the human
 CC lactadherin protein.

XX Sequence 387 AA;

Query Match 100.0%; Score 49; PR 21; Length 387;
 Best Local Similarity 100.0%; Pred. No. 6.42;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLLRRMWT 9
 Db 131 nllrrmwv 139

RESULT 4

AAM23507
 ID AAM23507 standard; Protein; 1763 AA.

XX AAM23507;

XX AC
 XX DT 12-OCT-2001 (first entry)

XX DE Human EST encoded protein SEQ ID NO: 1032.

XX KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
 KW diagnostics; forensic test; gene mapping; genetic disorder;
 KW biodiversity; gene therapy; nutrition.

XX OS Homo sapiens

XX PN W0200154477-A2.

XX XN 02-AUG-2001

XX PF 25-JAN 2001; 2001WO-0502687.

XX PP 25-JAN-2000; 2000US-0491404.

XX PR 17-JUL-2000; 2000US-0617746.

XX PR 03-AUG-2000; 2000US-0631451.

XX PR 15-SEP-2000; 2000US-0663870.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Zhao F, Chen XP, Wang J, Chen B, Asundi V,
 PI Cao Y, Drmanac RA, Zhang J, Wehrman T;

DR WPI: 2001-476144/51

DR N-PSDB: AAN98166.

XX PT Isolated polypeptide for treatment of diseases, diagnostics, raising
 PT antibodies and research use

XX PS Claim 20; Page 748 792, 1275pp; English.

XX CC The present invention provides the protein and coding sequences of novel
 CC proteins from a variety of organisms, including human, dog, cat, horse,
 CC cow, pig, hamster, monkey, macaque, yeast, bacterium, fruit fly, sea
 CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a
 CC protein of the invention.

XX Sequence 1763 AA;

Query Match 83.7%; Score 41; DP 22; Length 1763;
 Best Local Similarity 87.5%; Pred. No. 53;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLLRRMWT 8
 Db 217 nllrrmwv 224

RESULT 5

AAM78911

ID AAM78911 standard; Protein; 2473 AA.

XX AAM78911;

XX AC
 XX DT 06-NOV-2001 (first entry)

XX DE Human protein SEQ ID NO 1573.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
 KW disease growth factor; immunomodulatory activity; leukaemia;
 KW nervous system disorder; arthritis; inflammation.

XX OS Homo sapiens.

XX PN W0200157190-A2.

XX XN 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-0504098.

XX PP 03-FEB-2000; 2000US-0496914.

XX PR 27-APR-2000; 2000US-0569875.

XX PR 20-JUN-2000; 2000US-0598075.

XX PR 19-JUL-2000; 2000US-0620325.

XX PR 01-SEP-2000; 2000US-0654936.

XX PR 15-SEP-2000; 2000US-0663561.

XX PR 29-OCT-2000; 2000US-0693325.

XX PR 03-NOV-2000; 2000US-0728422.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RA, Asundi V, Chen F, Xu C, Cao Y, Ma Y,
 PI Zhao CA, Wang J, Wang J, Zhang J, Ren F, Chen F, Wang ZW,
 PI Xue AJ, Yang Y, Wehrman T, Goodrich P;

DR WPI: 2001-476283/51.

DR N-PSDB: AAK52044.

XX PT Nucleic acids encoding polypeptides with cytokine like activities,
 PT useful in diagnosis and gene therapy.

XX PS Claim 20; Page 3901-3905; 5231pp; English.

XX CC The invention relates to polypeptides (AAK51456-AAK54445) and the
 CC encoded polypeptides (AAM78911) that exhibit activity relating to
 CC cytokine, cell proliferation or cell differentiation of which may induce
 CC production of other cytokines in other cell populations. The
 CC polypeptides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine like activities,
 CC such as stem cell growth factor activity, hematopoiesis regulating
 CC activity, disease growth factor activity, immunomodulatory activity, and
 CC activity/inhibin activity, and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.

XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 4666
 CC (AAM78911) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.

XX Sequence 2473 AA;

Query Match 83.7%; Score 41; PR 22; Length 2473;
 Best Local Similarity 87.5%; Pred. No. 75;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLLRRMWV 8
 |||||:
 Db 927 NLLRRMWV 944

RESULT 6
 AAM79895
 ID AAM79895 standard; Protein; 2668 AA.

XX AC AAM79895;

XX DT 06-NOV-2001 (first entry)

XX DE Human protein SEQ ID NO 3541.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.

XX OS Homo sapiens.

XX PN W0200157190-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US04098.

XX PR 04-FEB-2000; 2000US-0496914.

XX PR 27-APR-2000; 2000US-0560875.

XX PR 26-JUN-2000; 2000US-0598075.

XX PR 19-JUL-2000; 2000US-0628326.

XX PR 01-SEP-2000; 2000US-0654936.

XX PR 15-SEP-2000; 2000US-0663571.

XX PR 29-OCT-2000; 2000US-0693325.

XX PR 30-NOV-2000; 2000US-0724222.

XX PA (BYSE) HYSEQ INC.

XX PI Tang YF, Liu C, Dhanase PT, Asundi V, Zhou P, Xu C, Gao Y, Mi Y;

PI Zhao QA, Wang D, Wang J, Zhang J, Fan F, Chen F, Wang ZW;

PI Xue AJ, Yang Y, Wehrman T, Goodrich P;

XX WP1: 2001-476283/51

XX WP2: AAK53828

XX PT Nucleic acids encoded polypeptides with cytokine like activities

XX PT useful in diagnosis and gene therapy -

XX PS Claim 20; page 377-479; 622pp; English.

XX CC The invention relates to polypeptides (AAK51456-AAK5493) and the

XX CC encoded polypeptides (AAK5423-AAK5437) that exhibit activity related to

XX CC cytokine, cell proliferation or cell differentiation, which may induce

XX CC production of other cytokines in other cell populations. The

XX CC polypeptides and polypeptides are useful in gene therapy, vaccines or

XX CC peptide therapy. The polypeptides have various cytokine-like activities,

XX CC e.g., stem cell growth factor activity, haematopoiesis regulating

XX CC activity, tissue growth factor activity, immunomodulatory activity and

XX CC activin inhibin activity and may be used in the diagnosis and/or

XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

XX CC inflammation

XX CC Notes: Records for SEQ ID NO 2116 (AAK52581), 2117 (AAK52582) and 2668

XX CC (AAK80020) are omitted as the relevant pages from the sequence listing

XX CC were missing at the time of publication

XX CC Sequence 2668 AA;

Query Match 83.7%; Score 41; PR 22; Length 2668;
 Best Local Similarity 87.5%; Pred. No. 81;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLLRRMWV 8
 |||||:
 Db 1122 NLLRRMWV 1129

RESULT 7
 AAU23593

ID AAU23593 standard; Protein; 112 AA.

XX AC AAU23593;

XX DT 18 DEC 2001 (first entry)

XX DE Novel human enzyme polypeptide #479.

XX KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;

XX KW ligase; hyperproliferative disorder; immunodeficiency disorder;

XX KW autoimmune disorder; neurological disorder; metabolic disorder;

XX KW inflammatory disorder; cardiovascular disorder; reproductive disorder;

XX KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;

XX KW nephrotropic; anticoagulant.

XX OS Homo sapiens.

XX PN W0200155301-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01239.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184544.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190326.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 09-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216647.

XX PR 17-JUL-2000; 2000US-0216880.

XX PR 11-SEP-2000; 2000US-0217487.

XX PR 11-SEP-2000; 2000US-0217496.

XX PR 14-SEP-2000; 2000US-0218290.

XX PR 29-SEP-2000; 2000US-0220963.

XX PR 29-JUL-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

XX PR 14-AUG-2000; 2000US-0225259.

XX PR 14-AUG-2000; 2000US-0225267.

XX PR 14-AUG-2000; 2000US-0225268.

XX PR 14-AUG-2000; 2000US-0225270.

XX PR 14-AUG-2000; 2000US-0225447.

XX PR 14-AUG-2000; 2000US-0225757.

XX PR 14-AUG-2000; 2000US-0225758.

XX PR 14-AUG-2000; 2000US-0225759.

XX PR 18-AUG-2000; 2000US-0226279.

XX PR 22-AUG-2000; 2000US-0226681.

XX PR 22-AUG-2000; 2000US-0226866.

XX PR 22-AUG-2000; 2000US-0227182.

XX PR 23-AUG-2000; 2000US-0227009.

XX PR 30-AUG-2000; 2000US-0228924.

XX PR 01-SEP-2000; 2000US-0229287.

XX PR 01-SEP-2000; 2000US-0229343.

XX PR 01-SEP-2000; 2000US-0229344.

ID ABG28776 standard; Protein; 666 AA.
 XX AC ABG28776;
 XX DT 18-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #28767.
 XX KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens.
 XX PN WC200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23 APR 2000; 2000US 0640167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX WI: 2001-639462/73.
 XX N-PSDB: AAS92963.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 20; SEQ ID No 54365; 103pp; English.
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligonucleotides, and for chromosome-
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (I) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detection of
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAG0010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pat_sequences.
 XX Sequence 666 AA:
 SQ
 Query Match 71.4%; Score 35; DB 22; Length 666;
 Best Local Similarity 75.0%; Pred. No. 2.5e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 NLLRRMW 8
 DB 137 nllrrmwi 144
 RESULT 9
 AAY51901
 ID AAY51901 standard; Protein; 329 AA.
 XX AC AAY51901;
 XX DT 15-JUN-2000 (first entry)
 XX DE Human 3GnT protein.
 XX KW UDP (uridine diphosphate)-N-acetylglucosaminyl; beta1,4-galactoside;
 KW beta1,3-N-acetylglucosaminyl transferase; 3GnT; polylactosaminyl;
 KW disaccharide production; oligosaccharide production; glycan;
 KW glycoprotein; gene therapy; transgenic animal; human.
 XX OS Homo sapiens.
 XX PN WO200011190-A2.
 XX PD 02-MAR-2000.
 XX PF 06-AUG-1999; 99WO-CH00365.
 XX PF 20-AUG 1998; 98CH-0001717.
 XX PA (BERG/) BERGER F G.
 XX PA (HENN/) HENNET T.
 XX PI Berger FG, Hennet T;
 XX WI: 2000-224787/19.
 XX N-PSDB: AAZ89455.
 XX New nucleic acid encoding a glucosaminyl transferase, used for in vivo
 PT or in vitro synthesis of glycan chains containing a specific
 PT disaccharide repeat element.
 XX Disclosure, Page 31-32, 34pp, German.
 XX This invention describes a novel isolated nucleic acid (I) encoding
 CC UDP (uridine diphosphate) N-acetylglucosaminyl:beta1,4-
 CC galactoside-beta1,3-N-acetylglucosaminyl transferase (3GnT)
 CC (polylactosaminyl type) 3GnTs are used, in vivo or in vitro, for
 CC enzymatic production of di- or oligo-saccharides, particularly glycans,
 CC containing the repeated disaccharide motif GlcNAc-beta1-3Gal. (I) is
 CC useful for recombinant expression of adals, to alter adhesion and
 CC recognition properties of transfected cells and to produce glycoproteins
 CC substituted by polylactosaminoglycans, when introduced into cells that
 CC express glycoproteins. Vectors containing (I) are used in gene therapy
 CC and for production of transgenic animals. Fragments of (I) are used to
 CC detect related sequences from other species and as antisense reagents.
 CC This sequence represents the human 3GnT protein described in the
 CC invention.
 XX Sequence 329 AA:
 SQ
 Query Match 69.4%; Score 34; DB 21; Length 329;
 Best Local Similarity 85.7%; Pred. No. 1.9e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 NLLRRMW 7
 DB 2 nllrraw 8
 RESULT 10
 AAU37812
 ID AAU37812 standard; Protein; 750 AA.
 XX AC AAU37812;
 XX DT 14-FEB-2002 (first entry)
 XX DE Streptococcus pneumoniae cellular proliferation protein #241.

KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.

XX
 XX 21-MAR-2001; 2001WO-US09180.
 XX
 XX 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-064918P.
 PR 26-MAY-2000; 2000US-067127P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253525P.
 PR 22-DEC-2000; 2000US-257921P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) FUTURE PHARM INC.
 XX
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Walli D, Trawick JD, Carr GJ;
 PI Yamamoto PT, Xu RH;
 XX
 DE WPI: 2001/611495/70
 DR N-PSDB: AAS55671.

XX New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 XX Example 3, Seq ID No 13405, 511pp, English.

XX The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pot_sequences

XX Sequence 750 AA;

Query Match 69.4%; Score 34; DB 22; Length 750.
 Best Local Similarity 55.6%; Pred. No. 4.3e-02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 NLIRRMWVT 9
 Db 109 sllqrftwt 117

RESULT 11
 ID ABG19088
 AC ABG19088 standard; Protein: 798 AA.
 XX
 AC ABG19088;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #19079.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US08631.
 XX
 XX 31-MAR-2000; 2000US-254021P.
 PR 23-AGO-2000; 2000US-064916P.
 XX
 XX (NYSE) HY3EQ INC.
 XX
 XX Brmanac RT, Liu C, Tang YL;
 XX
 XX WPI: 2001-629462/73.
 DR N-PSDB: AAS43275.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 XX Claim 20, Seq ID No 49447; 103pp, English.
 PS
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biochemical activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on RNA and
 CC saline acid sequences. ABG50016-ABG50377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pot_sequences.

XX Sequence 798 AA;

Query Match 69.4%; Score 34; DB 22; Length 798;
 Best Local Similarity 100.0%; Pred. No. 1.6e-02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 LLRRMW 7
 Db 362 LLRRMW 367

RESULT 12
 AAG55919
 ID AAG55919 standard; Protein: 187 AA.
 XX
 AC AAG55919;
 XX
 DT 11-FEB-2002 (first entry)
 XX
 DE Amino acid sequence of novel human arginine-rich protein (ARP).
 XX

KW ARP; angiogenesis; vascular endothelial growth factor; VEGF; cytostatic;
 KW arginine-rich protein; cardiant; antirheumatic; antiarthritic; human;
 KW antithrombotic; vasotropic; gynecological; antidiabetic; vulnerary;
 KW antitumor; dermatological; ophthalmological; antipsoriatic; apoptosis;
 KW gene therapy.

XX Homo sapiens.

XX Key location/Qualifiers

FT Misc-difference 16 /note "encoded by CCG"

FT Misc-difference 17 /note= "encoded by GTG"

FT Misc-difference 48 /note- "encoded by AAA"

FT Misc-difference 49 /note- "encoded by GAC"

FT Misc-difference 50..51

FT /note= "nucleotides encoding these 2 residues are not

FT indicated in the corresponding nucleotide

FT sequence"

XX W0200170174-A2.

PN 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09043.

XX 21-MAR-2000; 2000US 101201P

XX (CURA-) CURAGEN INC.

PA (GETH) GENENTECH INC.

XX Kastelli LR; Corbet H;

XX WPI; 2001-649087/73.

DR N-PSDB; AA167447.

XX Modulating angiogenesis and/or apoptosis for preventing or treating
 PT cancer, myocardial infarction and promoting healing, by modulating the
 PT activity of vascular endothelial growth factor-mediated gene
 PT polypeptide -

XX Claim 4; Page 15; 155pp; English.

XX The invention relates to modulating angiogenesis and cell survival that
 CC involves modulating the activity of at least one vascular endothelial
 CC growth factor (VEGF)-modulated gene polypeptide. The method is useful for
 CC modulating angiogenesis and cell survival, for treating tumour and cancer
 CC by decreasing angiogenesis in cancerous tumours and treating myocardial
 CC infarction and promoting healing, by increasing angiogenesis. Transgenic
 CC non-human animals, having disrupted arginine-rich protein (ARP), are
 CC useful for determining the clinical stage of ovarian tumours, which is
 CC useful for determining if the tumour has potential for metastasis. ARP is
 CC useful in gene therapy and in diagnostic applications. VEGFm proteins
 CC are useful in the treatment of tumours, neoplasias, hemangiomas,
 CC rheumatoid arthritis, atherosclerosis, idiopathic pulmonary fibrosis,
 CC vascular restenosis, arteriovenous malformations, melanoma, neurovascular
 CC glaucoma, psoriasis, hemophilic joints, hypertrophic scars, Oster-Wabet
 CC syndrome, scleroderma, vascular adhesion pathologies, synovitis,
 CC dermatitis, endometriosis, diabetic retinopathy, neovascularization
 CC associated with corneal injury or grafts, wound, sore, and ulcer healing.
 CC The present sequence represents the amino acid sequence of human ARP.

XX Sequence 187 AA;

Query Match 67 48; Score 33; Pp 22; Length 187;

Best Local Similarity 71.4%; Pred. No. 1.6e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 LRRMWT 9

|||||

Db 4 mrrmwat 10

RESULT 13

AAG28778

ID AAG28778 standard; Protein; 204 AA.

XX AC AAG28778;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID No: 34124.

XX KW Protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127452.

XX PR 06-APR-1999; 99US-0128244.

XX PR 08-APR-1999; 99US-0128714.

XX PR 16-APR-1999; 99US-0129845.

XX PR 19-APR-1999; 99US-0130077.

XX PR 21-APR-1999; 99US-0130449.

XX PR 23-APR-1999; 99US-0130510.

XX PR 23-APR-1999; 99US-0130891.

XX PR 28-APR-1999; 99US-0131449.

XX PR 30-APR-1999; 99US-0132048.

XX PR 30-APR-1999; 99US-0132407.

XX PR 04-MAY-1999; 99US-0132484.

XX PR 05-MAY-1999; 99US-0132485.

XX PR 06-MAY-1999; 99US-0132486.

XX PR 06-MAY-1999; 99US-0132487.

XX PR 07-MAY-1999; 99US-0132863.

XX PR 11-MAY-1999; 99US-0134256.

XX PR 14-MAY-1999; 99US-0134218.

XX PR 14-MAY-1999; 99US-0134219.

XX PR 14-MAY-1999; 99US-0134221.

XX PR 14-MAY-1999; 99US-0134370.

XX PR 16-MAY-1999; 99US-0134768.

XX PR 19-MAY-1999; 99US-0134941.

XX PR 20-MAY-1999; 99US-0135124.

XX PR 21-MAY-1999; 99US-0135453.

XX PR 24-MAY-1999; 99US-0135629.

XX PR 25-MAY-1999; 99US-0136021.

XX PR 27-MAY-1999; 99US-0136392.

XX PR 28-MAY-1999; 99US-0136782.

XX PR 01-JUN-1999; 99US-0137222.

XX PR 03-JUN-1999; 99US-0137528.

XX PR 04-JUN-1999; 99US-0137502.

XX PR 07-JUN-1999; 99US-0137724.

XX PR 08-JUN-1999; 99US-0138094.

XX PR 10-JUN-1999; 99US-0138540.

XX PR 14-JUN-1999; 99US-0138847.

XX PR 16-JUN-1999; 99US-0139119.

XX PR 16-JUN-1999; 99US-0139452.

XX PR 17-JUN-1999; 99US-0139453.

XX PR 17-JUN-1999; 99US-0139492.

XX PR 18-JUN-1999; 99US-0139454.

XX PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; qnrs-0139456;
PR 18-JUN-1999; qnrs-0139457;
PR 18-JUN-1999; qnrs-0139458;
PR 18-JUN-1999; qnrs-0139459;
PR 18-JUN-1999; qnrs-0139460;
PR 18-JUN-1999; qnrs-0139461;
PR 18-JUN-1999; qnrs-0139462;
PR 18-JUN-1999; qnrs-0139463;
PR 18-JUN-1999; qnrs-0139464;
PR 18-JUN-1999; qnrs-0139465;
PR 18-JUN-1999; qnrs-0139466;
PR 18-JUN-1999; qnrs-0139467;
PR 18-JUN-1999; qnrs-0139468;
PR 18-JUN-1999; qnrs-0139469;
PR 18-JUN-1999; qnrs-0139470;
PR 18-JUN-1999; qnrs-0139471;
PR 18-JUN-1999; qnrs-0139472;
PR 18-JUN-1999; qnrs-0139473;
PR 18-JUN-1999; qnrs-0139474;
PR 18-JUN-1999; qnrs-0139475;
PR 18-JUN-1999; qnrs-0139476;
PR 18-JUN-1999; qnrs-0139477;
PR 18-JUN-1999; qnrs-0139478;
PR 18-JUN-1999; qnrs-0139479;
PR 18-JUN-1999; qnrs-0139480;
PR 18-JUN-1999; qnrs-0139481;
PR 18-JUN-1999; qnrs-0139482;
PR 18-JUN-1999; qnrs-0139483;
PR 18-JUN-1999; qnrs-0139484;
PR 18-JUN-1999; qnrs-0139485;
PR 18-JUN-1999; qnrs-0139486;
PR 18-JUN-1999; qnrs-0139487;
PR 18-JUN-1999; qnrs-0139488;
PR 18-JUN-1999; qnrs-0139489;
PR 18-JUN-1999; qnrs-0139490;
PR 18-JUN-1999; qnrs-0139491;
PR 18-JUN-1999; qnrs-0139492;
PR 18-JUN-1999; qnrs-0139493;
PR 18-JUN-1999; qnrs-0139494;
PR 18-JUN-1999; qnrs-0139495;
PR 18-JUN-1999; qnrs-0139496;
PR 18-JUN-1999; qnrs-0139497;
PR 18-JUN-1999; qnrs-0139498;
PR 18-JUN-1999; qnrs-0139499;
PR 18-JUN-1999; qnrs-0139500;

PR 18-AUG-1999; qnrs-0149426;
PR 20-AUG-1999; qnrs-0149427;
PR 20-AUG-1999; qnrs-0149428;
PR 20-AUG-1999; qnrs-0149429;
PR 23-AUG-1999; qnrs-0149430;
PR 23-AUG-1999; qnrs-0149431;
PR 25-AUG-1999; qnrs-0150566;
PR 26-AUG-1999; qnrs-0150884;
PR 27-AUG-1999; qnrs-0151065;
PR 27-AUG-1999; qnrs-0151066;
PR 27-AUG-1999; qnrs-0151080;
PR 30-AUG-1999; qnrs-0151303;
PR 31-AUG-1999; qnrs-0151438;
PR 01-SEP-1999; qnrs-0151930;
PR 07-SEP-1999; qnrs-0152463;
PR 10-SEP-1999; qnrs-0153070;
PR 12-SEP-1999; qnrs-0153758;
PR 15-SEP-1999; qnrs-0154018;
PR 16-SEP-1999; qnrs-0154039;
PR 20-SEP-1999; qnrs-0154779;
PR 22-SEP-1999; qnrs-0155139;
PR 23-SEP-1999; qnrs-0155486;
PR 24-SEP-1999; qnrs-0155664;
PR 29-SEP-1999; qnrs-0156458;
PR 29-SEP-1999; qnrs-0156596;
PR 04-OCT-1999; qnrs-0157117;
PR 06-OCT-1999; qnrs-0157753;
PR 06-OCT-1999; qnrs-0157865;
PR 07-OCT-1999; qnrs-0158029;
PR 08-OCT-1999; qnrs-0158232;
PR 12-OCT-1999; qnrs-0158369;
PR 13-OCT-1999; qnrs-0159293;
PR 13-OCT-1999; qnrs-0159294;
PR 13-OCT-1999; qnrs-0159295;
PR 14-OCT-1999; qnrs-0159329;
PR 14-OCT-1999; qnrs-0159330;
PR 14-OCT-1999; qnrs-0159331;
PR 14-OCT-1999; qnrs-0159637;
PR 14-OCT-1999; qnrs-0159638;
PR 18-OCT-1999; qnrs-0159584;
PR 21-OCT-1999; qnrs-0160741;
PR 21-OCT-1999; qnrs-0160767;
PR 21-OCT-1999; qnrs-0160768;
PR 21-OCT-1999; qnrs-0160770;
PR 21-OCT-1999; qnrs-0160814;
PR 21-OCT-1999; qnrs-0160815;
PR 22-OCT-1999; qnrs-0160880;
PR 22-OCT-1999; qnrs-0160981;
PR 22-OCT-1999; qnrs-0160989;
PR 25-OCT-1999; qnrs-0161404;
PR 25-OCT-1999; qnrs-0161405;
PR 25-OCT-1999; qnrs-0161406;
PR 26-OCT-1999; qnrs-0161359;
PR 26-OCT-1999; qnrs-0161360;
PR 26-OCT-1999; qnrs-0161361;
PR 28-OCT-1999; qnrs-0161920;
PR 28-OCT-1999; qnrs-0161992;
PR 28-OCT-1999; qnrs-0161993;
PR 29-OCT-1999; qnrs-0162142;

Query Match 67.9; Score 39; DB 21; Length 204;
Best Local Similarity 62.5%; Pred. No. 186-02;
Matches 5; Conservative 2; Mismatches 1; Indels 2; Caps 0;

QY 1 NLRPMWV 8

DE 22 nvirrlwv 29

RESULT 14

AAB90765

ID AAB90765 standard; Protein; 234 AA.

XX

```

AC  AAB90765;
XX
DI  15-JUN-2001 (first entry)
DE  Human shear stress-response protein SEQ ID NO: 30.
XX
KW  Human shear stress-response protein; vascular disease;
KW  arteriosclerosis.
XX
OS  Homo sapiens.
XX
PN  W6200125427-A1
PD  12-APR-2001.
XX
PF  02-OCT-2000; 2000WG-IP06840
XX
PR  01-OCT-1999; 99JP-0280976.
XX
PA  (KYOW ) KYOWA HAKKO KOGYO KK.
PA  (NOJII) NOJIMA H.
XX
PI  Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada F,
PI  Fuga T, Sekine S, Nakamura Y, Sugano S.
XX
WP  2001:266308/27.
DR  N-PSDB; AAH02888.
XX
PT  DNA sequences, proteins encoded by them and antibodies against them
PT  useful in diagnosis and treatment of vascular disease caused by
PT  arteriosclerosis.
XX
PS  Claim 60; Page 258-259; 678pp; Japanese.
XX
CC  The present invention provides the protein and coding sequences of a
CC  number of human shear stress response proteins. These are useful in the
CC  diagnosis, treatment and screening of vascular diseases caused by
CC  arteriosclerosis, including heart failure, post-PTCA restenosis and
CC  hypertension.
XX
SQ  Sequence 234 AA;

Query Match: 67.3%; Score 32; DB 22; Length 234;
Best Local Similarity 71.4%; Pred. No. 20-02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY  3 LRRMMVT 9
DB  :||| 1
DB  53 mirrmat 59

RESULT 15
AA028777
ID  AAG28777 standard. Protein. 258 AA.
XX
AC  AAG28777;
XX
UT  17-OCT-2000 (first entry)
XX
DE  Arbidopsis thaliana protein fragment SEQ ID No. 34123.
XX
KW  Protein identification; signal transduction pathway; metabolic pathway;
KW  hybridisation assay; genetic mapping; gene expression analysis; promoter;
KW  termination sequence.
XX
OS  Arabidopsis thaliana.
XX
PN  EP1033405-A2.
XX
PD  06-SEP-2000.
XX
PF  25-FEB-2000; 2000EP-0301444

```

```

XX  25-FEB-1999; 99US-0121825.
XX  05-MAR-1999; 99US-0121180.
XX  09-MAR-1999; 99US-0123548.
XX  23-MAR-1999; 99US-0125788.
XX  25-MAR-1999; 99US-0126264.
XX  29-MAR-1999; 99US-0126785.
XX  01-APR-1999; 99US-0127462.
XX  06-APR-1999; 99US-0128234.
XX  08-APR-1999; 99US-0128714.
XX  16-APR-1999; 99US-0129845.
XX  19-APR-1999; 99US-0130077.
XX  21-APR-1999; 99US-0130449.
XX  23-APR-1999; 99US-0130510.
XX  23-APR-1999; 99US-0140891.
XX  28-APR-1999; 99US-0141449.
XX  30-APR-1999; 99US-0132048.
XX  30-APR-1999; 99US-0132407.
XX  04-MAY-1999; 99US-0132484.
XX  05-MAY-1999; 99US-0132485.
XX  06-MAY-1999; 99US-0132486.
XX  06-MAY-1999; 99US-0132487.
XX  07-MAY-1999; 99US-0132863.
XX  11-MAY-1999; 99US-0134256.
XX  14-MAY-1999; 99US-0134218.
XX  14-MAY-1999; 99US-0134219.
XX  14-MAY-1999; 99US-0134221.
XX  14-MAY-1999; 99US-0134370.
XX  18-MAY-1999; 99US-0134768.
XX  18-MAY-1999; 99US-0134941.
XX  20-MAY-1999; 99US-0135124.
XX  21-MAY-1999; 99US-0135353.
XX  24-MAY-1999; 99US-0135629.
XX  25-MAY-1999; 99US-0136021.
XX  27-MAY-1999; 99US-0136392.
XX  28-MAY-1999; 99US-0136782.
XX  01-JUN-1999; 99US-0137212.
XX  03-JUN-1999; 99US-0137528.
XX  04-JUN-1999; 99US-0137502.
XX  07-JUN-1999; 99US-0137724.
XX  08-JUN-1999; 99US-0138044.
XX  10-JUN-1999; 99US-0138540.
XX  10-JUN-1999; 99US-0138847.
XX  14-JUN-1999; 99US-0139119.
XX  16-JUN-1999; 99US-0139452.
XX  16-JUN-1999; 99US-0139453.
XX  17-JUN-1999; 99US-0139492.
XX  18-JUN-1999; 99US-0139454.
XX  18-JUN-1999; 99US-0139455.
XX  18-JUN-1999; 99US-0139456.
XX  18-JUN-1999; 99US-0139457.
XX  18-JUN-1999; 99US-0139458.
XX  18-JUN-1999; 99US-0139459.
XX  18-JUN-1999; 99US-0139460.
XX  18-JUN-1999; 99US-0139461.
XX  18-JUN-1999; 99US-0139462.
XX  18-JUN-1999; 99US-0139463.
XX  18-JUN-1999; 99US-0139750.
XX  18-JUN-1999; 99US-0139763.
XX  21-JUN-1999; 99US-0139817.
XX  22-JUN-1999; 99US-0139899.
XX  23-JUN-1999; 99US-0140353.
XX  23-JUN-1999; 99US-0140354.
XX  24-JUN-1999; 99US-0140695.
XX  28-JUN-1999; 99US-0140823.
XX  29-JUN-1999; 99US-0140991.
XX  30-JUN-1999; 99US-0141287.
XX  01-JUL-1999; 99US-0141842.
XX  01-JUL-1999; 99US-0142154.
XX  02-JUL-1999; 99US-0142055.
XX  06-JUL-1999; 99US-0142390.
XX  08-JUL-1999; 99US-0142803.
XX  09-JUL-1999; 99US-0142900.

```



```

Matches      9;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

QY      1 NLEETPILA 9
      11111111
Db      128 NLEETPILA 136

RESULT      2
US-07-607-538C-3
; Sequence 3, Application US/07607538C
; Patent No. 5455031
; GENERAL INFORMATION:
; APPLICANT: Ceriani Dr., Roberto L.
; APPLICANT: Peterson Dr., Jerry A.
; APPLICANT: Larocca, David J.
; TITLE OF INVENTION: DIFFERENTIATION ANTIGEN BINDING SPECIFICITY AND CLOTTING
; TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
; TITLE OF INVENTION: FUSION PROTEIN, POLYNUCLEOTIDE AND POLYPEPTIDE,
; TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-
; TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: V. Amzel & Assoc.
; STREET: 2055 No. 5455031th Broadway
; CITY: Walnut Creek
; STATE: California
; COUNTRY: USA
; ZIP: 94596
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS 5.0
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/607,538C
; FILING DATE: 01-NOV-1990
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Viviana Anzel
; REGISTRATION NUMBER: 30,930
; REPRESENTATION NUMBER: 30930-004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 943-1193
; TELEFAX: (510) 943-1189
; TELEX: N.A.
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 217 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE:
US-07-607-538C-3

Query Match      100.0%; Score 45; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches      9;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

QY      1 NLEETPILA 9
      11111111
Db      186 NLEETPILA 194

RESULT      4
US-08-162-402B-3
; Sequence 3, Application US/08162402B
; Patent No. 5972447
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.

```

```

; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LARocca, DAVID J.
; TITLE OF INVENTION: 45 KILTAIN HEPAN MILK FAT
; TITLE OF INVENTION: GLYCOPOLY (HMPG) ANTIGEN, FRAGMENTS & PEPTON PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: 03-08-042,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 217 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-162-402B-3

Query Match      100.0%; Score 45; DB 2; Length 217;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches      9;  Conservative      0;  Mismatches      0;  Indels      0;  Gaps      0;

QY      1 NLEETPILA 9
      11111111
Db      186 NLEETPILA 194

RESULT      4
US-07-607-538C-2
; Sequence 2, Application US/07607538C
; Patent No. 5455031
; GENERAL INFORMATION:
; APPLICANT: Ceriani Dr., Roberto L.
; APPLICANT: Peterson Dr., Jerry A.
; APPLICANT: Larocca, David J.
; TITLE OF INVENTION: POLYPEPTIDE WITH 46
; TITLE OF INVENTION: DIFFERENTIATION ANTIGEN BINDING SPECIFICITY AND CLOTTING
; TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
; TITLE OF INVENTION: FUSION PROTEIN, POLYNUCLEOTIDE AND POLYPEPTIDE,
; TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-
; TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: V. Amzel & Assoc.
; STREET: 2055 No. 5455031th Broadway
; CITY: Walnut Creek
; STATE: California
; COUNTRY: USA

```


ZIP: 94506
COMPUTER PERIPHERAL FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 5.0
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/607,538C
FILING DATE: 01-NOV-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Viviana Amzel
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: P667-054
TELEPHONE: (510) 943-1931
TELEFAX: (510) 943-1189
TELEX: N.A.
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE:
US-07-607-538C-2

Query Match 100.0%; Score 45; DB 1; Length 218;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPILA 9
Db 187 NLFETPILA 195

RESULT 5
US-08-162-402B-2
Sequence 2, Application US/08162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERTANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LARGOCCA, DAVID J.
TITLE OF INVENTION: 46 KILTON HUMAN MILK FAT
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplawski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162,402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: P66 38215
TELECOMMUNICATION INFORMATION:

TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-2

Query Match 100.0%; Score 45; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 0.055;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPILA 9
Db 187 NLFETPILA 195

RESULT 6
US-08-162-402B-6
Sequence 6, Application US/08162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERTANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LARGOCCA, DAVID J.
TITLE OF INVENTION: 46 KILTON HUMAN MILK FAT
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplawski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162,402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: P66 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 387 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-6

Query Match 100.0%; Score 45; DB 2; Length 387;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLETPHIA 9
 |||||
 DB 456 NLETPHIA 364

RESULT 7

US-08-162-402B-8
 : Sequence 8, Application US/08162402B
 : Patent No. 5972337
 : GENERAL INFORMATION:
 : APPLICANT: CERIANT, ROBERTO L.
 : APPLICANT: PETERSON, JERRY A.
 : APPLICANT: LAROCCA, DAVID J.
 : TITLE OF INVENTION: 46 KLALTON HUMAN MILK FAT
 : NUMBER OF SEQUENCES: 29
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Pretty, Schroeder & Poplowski
 : STREET: 444 South Flower St., 19th Floor
 : CITY: Los Angeles
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 90071
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette
 : COMPUTER: IBM Compatible
 : OPERATING SYSTEM: DOS
 : SOFTWARE: FastSeq for Windows Version 2.0
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/162.402B
 : FILING DATE: 03-DEC-1993
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER:
 : FILING DATE:
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Amzel, Viviana
 : REGISTRATION NUMBER: 30,930
 : REFERENCE/DOCKET NUMBER: P66 38215
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 213-622-7700
 : TELEFAX: 213-489-4210
 : TELEX:
 : INFORMATION FOR SEQ ID NO: 7:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 465 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: unknown
 : TOPOLOGY: unknown
 : MOLECULE TYPE: peptide
 : US-08-162-402B-8

Query Match 100.0%; Score 45; DB 2; Length 465;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 1 NLETPHIA 9
 |||||
 DB 434 NLETPHIA 442

RESULT 8

US-08-162-402B-10
 : Sequence 10, Application US/08162402B
 : Patent No. 5972337
 : GENERAL INFORMATION:
 : APPLICANT: CERIANT, ROBERTO L.
 : APPLICANT: PETERSON, JERRY A.
 : APPLICANT: LAROCCA, DAVID J.
 : TITLE OF INVENTION: 46 KLALTON HUMAN MILK FAT
 : NUMBER OF SEQUENCES: 29
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Pretty, Schroeder & Poplowski
 : STREET: 444 South Flower St., 19th Floor
 : CITY: Los Angeles
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 90071
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette
 : COMPUTER: IBM Compatible
 : OPERATING SYSTEM: DOS
 : SOFTWARE: FastSeq for Windows Version 2.0
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/162.402B
 : FILING DATE: 03-DEC-1993

: NUMBER OF SEQUENCES: 29
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Pretty, Schroeder & Poplowski
 : STREET: 444 South Flower St., 19th Floor
 : CITY: Los Angeles
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 90071
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette
 : COMPUTER: IBM Compatible
 : OPERATING SYSTEM: DOS
 : SOFTWARE: FastSeq for Windows Version 2.0
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/162.402B
 : FILING DATE: 03-DEC-1993
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER:
 : FILING DATE:
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Amzel, Viviana
 : REGISTRATION NUMBER: 30,930
 : REFERENCE/DOCKET NUMBER: P66 38215
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 213-622-7700
 : TELEFAX: 213-489-4210
 : TELEX:
 : INFORMATION FOR SEQ ID NO: 10:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 160 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: unknown
 : TOPOLOGY: unknown
 : MOLECULE TYPE: peptide
 : US-08-162-402B-10

Query Match 82.3%; Score 27; DB 2; Length 160;
 Best Local Similarity 77.8%; Pred. No. 1.8;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLETPHIA 9
 |||||
 DB 126 NLETPHIA 134

RESULT 9

US-08-162-402B-26
 : Sequence 26, Application US/08162402B
 : Patent No. 5972337
 : GENERAL INFORMATION:
 : APPLICANT: CERIANT, ROBERTO L.
 : APPLICANT: PETERSON, JERRY A.
 : APPLICANT: LAROCCA, DAVID J.
 : TITLE OF INVENTION: 46 KLALTON HUMAN MILK FAT
 : NUMBER OF SEQUENCES: 29
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Pretty, Schroeder & Poplowski
 : STREET: 444 South Flower St., 19th Floor
 : CITY: Los Angeles
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 90071
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette
 : COMPUTER: IBM Compatible
 : OPERATING SYSTEM: DOS
 : SOFTWARE: FastSeq for Windows Version 2.0
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/162.402B
 : FILING DATE: 03-DEC-1993


```

; APPLICANT: Perou, Charles
; APPLICANT: Moore, Karen
; TITLE OF INVENTION: COMPOSITIONS FOR THE DIAGNOSIS
; AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/396,540
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/822,445
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-062-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3672 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-09-396-540-12

```

```

Query Match 71.1%; Score 32; DB 4; Length 3672;
Best Local Similarity 75.0%; Pred. No. 7.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 5; Gaps 0.

```

```

Qy 1 NLFETPIL 8
| | | | | : |
Db 443 NLFETAVL 450

```

```

RESULT 13
; Sequence 10, Application US/09/396,540
; Patent No. 5,952,223
; GENERAL INFORMATION:
; APPLICANT: Kaplan, Jerry
; APPLICANT: Perou, Charles
; APPLICANT: Moore, Karen
; TITLE OF INVENTION: COMPOSITIONS FOR THE DIAGNOSIS
; AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/822,445
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-062-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3801 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-822-445-10

```

```

Query Match 71.1%; Score 32; DB 2; Length 3801;
Best Local Similarity 75.0%; Pred. No. 8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0.

```

```

Qy 1 NLFETPIL 8
| | | | | : |
Db 443 NLFETAVL 450

```

```

RESULT 14
; Sequence 10, Application US/09396540
; Patent No. 6310182
; GENERAL INFORMATION:
; APPLICANT: Kaplan, Jerry
; APPLICANT: Perou, Charles
; APPLICANT: Moore, Karen
; TITLE OF INVENTION: COMPOSITIONS FOR THE DIAGNOSIS
; AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/396,540
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/822,445
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-062-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3801 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown

```

; MOLECULE TYPE: protein
US-09-396-340-10

Query Match 71.1% Score 32 DB 4 Length 3801
Best Local Similarity 75.04; Pred. No. 8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 1; Gaps 0;

QY 1 NLFETPIL 8
DB 443 NLFETAVL 450

RESULT 15

US-08-480-229C-2
; Sequence 2, Application US/08480229C
; Patent No. 5874562
; GENERAL INFORMATION:
; APPLICANT: Quettermous, Thomas
; APPLICANT: Hogan, Brigid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; TITLE OF INVENTION: CELL LOCUS-1
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-POS/MS-POS
; SOFTWARE: Patent In Release #1 0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08480,229C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0026-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 85 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-480-229C-2

Query Match 68.98; Score 31; DB 2; Length 85;
Best Local Similarity 55.6%; Pred. No. 15;
Matches 5; Conservative 2; Mismatches 2; Indels 2; Gaps 0;

QY 1 NLFETPILA 9
DB 67 NLFETPMA 75

Search completed: September 5, 2002, 15:27:34
Job time: 319 sec

XX WP1; 2000-205463/18.
 XX
 PT Tumor associated antigen peptides, especially derived from uroplakin,
 PT useful as vaccines to prevent or cure cancers including breast,
 PT bladder, prostate, pancreas, ovary, thyroid, colon and stomach
 PS Claim 17; Page 99; 11pp; English.
 XX
 CC Tumor associated antigen peptides (TAA) may be used for the
 CC treatment, prevention and cure of cancer or cancer metastases. The
 CC cancer may be breast, bladder, prostate, pancreas, ovary, thyroid,
 CC colon, stomach, head or neck cancer or a carcinoma. The tumor
 CC associated antigens are presentable to the immune system by HLA-A2
 CC molecules and are generally between 8 to 10 amino acids in length.
 CC The amino acids located at positions 2 and 9 of the tumor associated
 CC antigens are the anchor residues which participate in the binding to
 CC MHC class I molecules, more specifically HLA-A2. More tumor
 CC associated antigens are described in GENESQ records AAY82806-Y82882.
 CC Those tumor associated antigens described in records AAY82836-Y82854
 CC and AAY82855-Y82869 are derived from uroplakin, such as Uroplakin II,
 CC Uroplakin Ia, Uroplakin III and Uroplakin Ib. Those described in
 CC records AAY82825-Y82829 are derived from prostate specific antigen
 CC (PSA). Those described in records AAY82830-Y82835 are derived from
 CC prostate specific membrane antigen (PSMA). Those described in
 CC records Y82846-AAY82849 are derived from prostate acid phosphatase
 CC (PAP). Those described in records AAY82940-Y82946 are derived
 CC from lactadherin (BA 46). Those described in records AAY82847-Y82854
 CC are derived from Macin and those described in records AAY82871-Y82882
 CC are derived from Teratocarcinoma derived growth factor (CRPTO-1).
 XX
 XX Sequence 9 AA;
 SQ

Query Match 100.0%; Score 45; IP 21; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e-65;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q7 1 NLEETPILA 9
 LB 1 Nletpila 9
 RESULT 2
 AAY77253
 ID AAY77253 standard; Protein: 217 AA.
 AC AAY77253;
 XX
 XX 21-NOV-1995 (first entry)
 XX
 XX HMEG 46 kDa antigen C-terminal region.
 XX
 KW HMEG; human milk fat globule; antigen; immunogen; vaccine; virucide;
 KW epithelium; tumor; breast cancer; monoclonal antibody; MAB.
 XX
 OS Homo sapiens.
 XX
 PN W09515171-A.
 XX
 PD 08-JUN-1995.
 XX
 PF 05-DEC-1994; 94WO-0813467.
 XX
 PR 03-DEC-1993; 93US-0162402.
 XX
 PA (CANC-) CANCER RES FUND CONTRA COSTA.
 XX
 PI Ceriani RL, Larocca DL, Peterson JA;
 XX
 DR WP1; 1995-215151/28.
 XX
 PR 03-DEC-1993; 93US-0162402.
 XX
 PA (CANC-) CANCER RES FUND CONTRA COSTA.
 XX
 PI Ceriani RL, Larocca DL, Peterson JA;
 XX
 DR WP1; 1995-215151/28.
 XX
 PT 46 kD apparent molecular weight human milk fat globule antigen -

PT used in assays to determine the presence of a cancerous tumour of
 PT epithelial origin, and in a vaccine against neoplastic tumours
 XX
 PS Claim 7; Page 41; 68pp; English.
 XX
 CC A partial cDNA clone BA46.1 (AA051155) for the 46 kDa HMEG antigen,
 CC a major component of the apical surface of the normal breast
 CC epithelial cell, was obtained by PCR of cDNA from a lactating
 CC breast cDNA library. The C-terminal region of the encoded protein
 CC showed 43% identity to corresp. regions of human Factor-V and 48%
 CC to Factor-VIII.
 XX
 SQ Sequence 217 AA;
 SQ

Query Match 100.0%; Score 45; IP 16; Length 217;
 Best Local Similarity 100.0%; Pred. No. 0.37;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q7 1 NLEETPILA 9
 LB 1 Nletpila 194
 RESULT 3
 AAY77254
 ID AAY77254 standard; Protein: 218 AA.
 XX
 AC AAY77254;
 XX
 XX 21-NOV-1995 (first entry)
 XX
 XX HMEG 46 kDa antigen partial sequence.
 XX
 KW HMEG; human milk fat globule; antigen; immunogen; vaccine; virucide;
 KW epithelium; tumor; breast cancer; monoclonal antibody; MAB.
 XX
 OS Homo sapiens.
 XX
 PN W09515171-A.
 XX
 PD 08-JUN-1995.
 XX
 PF 05-DEC-1994; 94WO-0813467.
 XX
 PR 03-DEC-1993; 93US-0162402.
 XX
 PA (CANC-) CANCER RES FUND CONTRA COSTA.
 XX
 PI Ceriani RL, Larocca DL, Peterson JA;
 XX
 DR WP1; 1995-215151/28.
 XX
 PR N-PSDR; AA051155.
 XX
 PT 46 kD apparent molecular weight human milk fat globule antigen -
 PT used in assays to determine the presence of a cancerous tumour of
 PT epithelial origin, and in a vaccine against neoplastic tumours
 XX
 PS Example 7; Page 38-40; 68pp; English.
 XX
 CC A partial cDNA clone BA46.1 (AA051155) for the 46 kDa HMEG antigen,
 CC a major component of the apical surface of the normal breast
 CC epithelial cell, was obtained by PCR of cDNA from a lactating

CC breast cDNA library.
XX
SQ Sequence 218 AA;

Query Match 100.0%; Score 45; DB 16; Length 218;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLFETPILA 9
|||||
Db 187 nlfetpila 195

RESULT 4

AAR77252
ID AAR77252 standard; Protein, 387 AA.

XX
AC AAR77252;

XX
DI 21-NOV-1995 (first entry)

XX
DE HMFG 46 kDa antigen.

XX
KW HMFG; human milk fat globule; antigen; immunogen; vaccine; virucide;
epithelium; tumor; breast cancer; monoclonal antibody; MAb.

XX
OS Homo sapiens.

XX
PN W09515171-A.

XX
PD 08-JUN-1995.

XX
PF 05-DEC-1994; 94WO-US13967.

XX
PP 03-DEC-1993; 93US-0162402.

XX
PA (CANC-) CANCER RES FUND CONTRA COSTA.

XX
PI Ceriani RL, Larocca DJ, Fetselson JA.

XX
DR WPI; 1995 215153/28.

XX
DR N-PSDB; AA091168.

XX
PI 46 kD apparent molecular weight human milk fat globule antigen
used in assays to determine the presence of a cancerous tumour of
epithelial origin, and in a vaccine against neoplastic tumours

XX
PS Claim 6: Page 46-47; 68pp; English

XX
CC A complete cDNA sequence for the 46 kDa HMFG antigen, a major
component of the apical surface of the normal breast epithelial
cell, was obtained by PCR and RACE methods. cDNA clones can be
used to prepare MAbs for use in immunotherapy, immunohistopathology,
prognosis, diagnosis, imaging and therapy. Recombinant antigen can
be expressed in prokaryotic or (glycosylated) in eukaryotic cells.

XX
SQ Sequence 387 AA;

Query Match 100.0%; Score 45; DB 16; Length 387;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLFETPILA 9
|||||
Db 356 nlfetpila 364

RESULT 5

AA94453

ID AAY94453 standard; Protein; 387 AA.

XX

AC AAY94453;
XX
DT 11-SEP-2000 (first entry)
XX
DE Human lactadherin protein.
XX
KW Human; lactadherin; MGF-E8; anti-tumour; immune response;
exosome; dendritic cell.
XX
OS Homo sapiens.

XX
FH Key location/qualifiers
FT Peptide 1..23
FT /label Secretion_signal
FT Protein 24..387
FT /label lactadherin
FT Binding-site 46..48
FT /label Integrin_binding_site

XX
PN EP1004664-A1.

XX
PD 31-MAY-2000.

XX
PF 24-NOV-1998; 98EP-0402925.

XX
PP 24-NOV-1998; 98EP-0402925.

XX
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
(CURT-) INST CURTE.

XX
DR WPI; 2000-352597/31.

XX
DR N-PSDB; AAA27140.

XX
PT Chimeric isolated (human) lactadherin polypeptide that functions as an
adaptor of cross priming to eliminate pathogenic antigens, e.g. in
tumors.

XX
PS Example 3; Page 12; 20pp; English.

XX
CC Lactadherin protein was found in exosomes produced by dendritic cells.
The protein is involved in the phagocytosis of particulate antigens by
dendritic cells. Exosomes produced by dendritic cells exposed to
tumour antigens induce potent immune responses. Lactadherin or variants
of it may be used in the mediation of an immune response. Variants of
lactadherin may be used for inhibition and/or stimulation of the
cross priming of antigens and stimulation of the phagocytosis of
antigens by dendritic cells. Compositions derived from lactadherin can
also be used to monitor an immune response, more specifically a CTL
(cytotoxic T-lymphocyte) response and also to produce CTLs specific
for a selected antigen. The present sequence is the human
lactadherin protein.

XX
SQ Sequence 387 AA;

Query Match 100.0%; Score 45; DB 17; Length 387;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 NLFETPILA 9

Db 356 nlfetpila 364

RESULT 6

AA92843

ID AAY82843 standard; peptide; 9 AA.

XX

AC AAY82843;

XX

DT 19-JUN-2000 (first entry)

XX

DE Lactadherin (HA-46) peptide fragment (tumour associated antigen).

XX Tumor associated antigen peptide; TAA; cancer; carcinoma;
 KW treatment; prevention; cure; anti-tumour vaccine; metastases;
 KW breast; bladder; prostate; pancreas; ovary; thyroid; colon;
 KW stomach; carcinoma; MHC Class I; HLA-A2; human;
 KW Major Histocompatibility Complex; uroplakin;
 KW prostate specific antigen; prostate specific membrane antigen;
 KW prostate acid phosphatase; mucin; lactadherin;
 KW teratocarcinoma derived growth factor; PSA; PSMA; PAP; CRYPTO-1
 OS Homo sapiens.
 XX WO200006723-A1.
 XX 10-FEB-2000.
 XX 24-MAR-1999; 96WO-1106417.
 XX 30-JUL-1998; 98IL-0125608.
 XX (YEDA) YEDA RES & DEV CO LTD.
 PA (BIOT-) BIO-TECHNOLOGY GEN CORP.
 XX Eisenbach L, Carmon L, Tirosh B, Bar-haim E, Paz A, Fridkin M;
 PI Fitzner-atlas C;
 XX WPI: 2000-205463/18.
 XX Tumor associated antigen peptides, especially derived from uroplakin,
 PI useful as vaccines to prevent or cure cancers including breast,
 PI bladder, prostate, pancreas, ovary, thyroid, colon and stomach .
 XX Claim 17; Page 100; 113pp; English.
 XX Tumor associated antigen peptides (TAA) may be used for the
 CC treatment, prevention and cure of cancer or cancer metastases. The
 CC cancer may be breast, bladder, prostate, pancreas, ovary, thyroid,
 CC colon, stomach, head or neck cancer or a carcinoma. The tumour
 CC associated antigens are presentable to a carcinoma. The tumour
 CC molecules and are generally between 8 to 10 amino acids in length.
 CC The amino acids located at positions 2 and 9 of the tumour associated
 CC antigens are the anchor residues which participate in the binding to
 CC MHC class I molecules, more specifically HLA-A2. More tumour
 CC associated antigens are described in GENESEQ records AAY82806-Y82882.
 CC Those tumour associated antigens described in records AAY82806-Y82834
 CC and AAY82855-Y82869 are derived from uroplakin, such as uroplakin II,
 CC uroplakin IV, uroplakin III and uroplakin IB. Those described in
 CC records AAY82825-Y82829 are derived from prostate specific antigen
 CC (PSA). Those described in records AAY82840-Y82846 are derived from
 CC prostate specific membrane antigen (PSMA). Those described in
 CC records Y82846 AAY82839 are derived from prostate acid phosphatase
 CC (PAP). Those described in records AAY82840-Y82846 are derived
 CC from lactadherin (LA-46). Those described in records AAY82847-Y82854
 CC are derived from Mucin and those described in records AAY82871-Y82882
 CC are derived from Teratocarcinoma derived growth factor (CRYPTO-1).
 XX SQ Sequence 9 AA;
 Query Match 82.2%; Score 37; DB 21; Length 9;
 Best Local Similarity 77.8%; Pred. No. 6.4e+05;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 NIFETPILA 9
 DB 1 nifetpila 9
 RESULT 7
 AAB62713
 ID AAB62713 standard; Protein: 1284 AA.
 XX JP05244959-A.
 AC AAB62713;

XX 26-MAR-2002 (first entry)
 DT Brosophila melanogaster polypeptide SEQ ID NO 14931.
 XX Brosophila melanogaster polypeptide; cell signalling; insecticide;
 KW Brosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX Brosophila melanogaster
 OS WO200171042-A2.
 XX 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US05231.
 PF 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 PA (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI: 2001-655869/75.
 DR N-PSDR; ABL06816.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PI genes from Brosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX Disclosure, SEQ ID NO 14931, 21pp - Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Brosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (AB01840 ABL16175) and the encoded proteins
 CC (AB057237-AB072072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pat_sequences.
 XX SQ Sequence 1284 AA;
 Query Match 77.8%; Score 35; DB 22; Length 1284;
 Best Local Similarity 46.7%; Pred. No. 2.4e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 NIFETPILA 9
 DB 154 nifetpila 162
 RESULT 8
 AAR42839
 ID AAR42839 standard; Protein: 1835 AA.
 XX AAR42839;
 AC AAR42839;
 DT 10-MAY-1994 (first entry)
 DE Urea amidolyase.
 XX Urea amidolyase.
 KW Urea amidolyase; URL; yeast; recombinant plasmid.
 XX Saccharomyces cerevisiae.
 OS JP05244959-A.
 PN JP05244959-A.
 XX 24-SEP-1993.
 PD

XX 05-MAR-1992; 92JP-0084541.
 XX 06-MAR-1992; 92JP-0084541.
 XX (TOYM) TOYOBO KK.
 XX WPI: 1993-138925/43
 DR N-PSDB: AA049460.
 XX
 XX DNA having the genetic information of urea amidolyase originated
 PT from Saccharomyces yeast can be used to prepare high purity
 PT urea amidolyase by culturing the transformant comprising the DNA
 XX
 XX Claim 2, Page 10-17, 17pp, Japanese.
 XX This sequence represents a protein which has urea amidolyase (UreA)
 CC activity and is derived from yeast. The DNA encoding this protein
 CC may be used within a recombinant plasmid for the production of highly
 CC pure URE.
 XX
 XX Sequence 1835 AA;
 SQ
 Query Match 77.8%; Score 35; DB 14; Length 1835;
 Best Local Similarity 100.0%; Pred. No. 35002;
 Matches 7; Conservative 3; Mismatches 0; Indels 3; Caps 3;
 QY 3 PEPPIIA 9
 Db 1418 tetpilla 1424
 |||||
 RESULT 9
 AAB41545
 ID AAB41545 standard; Protein: 171 AA.
 XX
 AC AAB41545;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF1309 polypeptide sequence SEQ ID NO.2618.
 XX
 KW Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; estrogenic; antithrombotic; immunosuppressant; cardiac;
 KW immunostimulant; thrombolytic; coagulant; vasoregic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 XX WC2000059473-A2.
 PN
 XX
 XX 05-OCT-2000.
 PD
 XX
 XX 31-MAR-2000; 2000W0-US06621.
 PF
 XX
 XX 31-MAR-1999; 99US-0127607.
 PR
 XX 02-APR-1999; 99US-0127636
 PR
 XX 05-APR-1999; 99US-0127729
 PR
 XX 30-MAR-2000; 2000US-0540763
 PR
 XX
 XX (CURA-) CURAGEN CORP.
 XX

PI Shimkets RA, Leach M;
 XX
 DR WPI: 2000 60262/57.
 DR N-PSDB: AAC75754.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating c.c. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease
 XX
 PS Claim 11; Page 1863 1864; 5507pp, English.
 XX
 XX AAC74435 to AAC77609 encode the proteins given in AAC40237 to AAC43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic, hepatotropic, vulnery;
 CC antiproliferative, antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antithrombotic; coagulant; vasoregic;
 CC immunostimulant; cardiac; thrombolytic; immunosuppressant; antidiabetic;
 CC antianemic; hypothyroid; dermatological; antirheumatic; antithyroid;
 CC antinflammatory; antibacterial; antifungal; antirheumatic; antithyroid;
 CC antianemic; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 171 AA;
 Query Match 75.6%; Score 34; DB 21; Length 171;
 Best Local Similarity 66.7%; Pred. No. 44;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Caps 0;
 QY 1 NLEPPIIA 9
 Db 27 ninepills 35
 |||||
 RESULT 10
 AAB23032
 ID AAB23032 standard; Protein: 196 AA.
 XX
 AC AAB23032;
 XX
 DT 15-JAN-2001 (first entry)
 XX
 DE Human putative mitochondrial protein, SREX 2982339.
 XX
 KW SEX protein, human, secreted, membrane-associated; cancer;
 KW proliferation regulator; differentiation regulator; non-malignant tumour;
 KW immune disorder; autoimmune disease; transplant rejection; allergy; AIDS;
 KW infection; inflammatory disorder; arthritis; haematopoietic disorder;
 KW skin disorder; cardiovascular disorder; atherosclerosis; testicular;
 KW neurological disease; Alzheimer's disease; trauma; wound;
 KW spinal cord injury; skeletal disorder; cytostatic; immunosuppressive;
 KW anti-HIV; antinflammatory; antithrombotic; antidiabetic;
 KW neuroprotective; vulnery; antiallergic; antimicrobial; cardiac;
 KW dermatological; gene therapy.
 XX
 OS Homo sapiens.
 XX
 XX WC2000053742 A2.
 PN
 XX
 XX 14-SEP-2000.
 PD
 XX
 XX 09-MAR-2000; 2000W0-US06280.
 PF

XX 09-MAR-1999; 940S-0124667.
 PR 08-MAR-2000; 2000US-0124667.
 XX (CURA-3) CURAGEN CORP.
 XX Shinketsu PA;
 XX
 XX WPI; 2000-594118/56
 DR N-PSDB; AAA94619.
 XX Novel human membrane associated or secreted polypeptides and
 PI polynucleotides useful for diagnosis, prevention and treatment of
 PT pathological states such as cancer, immune, cardiovascular and
 PT neurological disorders.
 XX
 XX Claim 1; Fig 4; 151pp; English.
 XX
 XX Sequence AAB3025 R23048 represent human SECX proteins. The SECX
 CC proteins of the invention are either secreted or membrane associated
 CC proteins and act as regulator of cellular proliferation and
 CC differentiation. SECX proteins or nucleotides are useful for diagnosing
 CC the presence of, or predisposition to, a disease associated with altered
 CC levels of SECX proteins and nucleotides. The SECX proteins are also
 CC useful to screen compounds that modulate SECX activity or expression. The
 CC interaction of a SECX protein with other cellular proteins may be useful
 CC to modulate the activity of a partner protein, cellular proliferation,
 CC cellular differentiation and cell survival. SECX nucleotides are useful
 CC for the recombinant expression of SECX protein, and may be used to detect
 CC SECX mRNA or genetic lesions in the SECX gene. They may also be used to
 CC modulate SECX expression (e.g., using antisense oligonucleotides). SECX
 CC nucleic acid sequences are also useful for identifying a cell or tissue
 CC type in a biological sample, and in forensic biology. SECX primers or
 CC probes are useful for detecting the presence of SECX nucleotides and for
 CC screening tissue cultures for contamination. Diseases that may be treated
 CC or prevented using SECX proteins or nucleotides include cancer (e.g.,
 CC colorectal carcinoma, prostate cancer), benign tumours, immune disorders
 CC (including autoimmune diseases, transplant rejection, allergies, AIDS),
 CC infections, inflammatory disorders, arthritis, haematopoietic disorders,
 CC skin disorders, cardiovascular disorders, attherosclerosis, osteoporosis,
 CC neurological diseases (e.g., Alzheimer's disease), trauma (e.g.,
 CC surgical or traumatic wounds, spinal cord injury), and skeletal
 CC disorders.
 XX
 XX Sequence 196 AA;
 SQ
 Query Match 75.6%; Score 34; DB 21; Length 196;
 Best Local Similarity 75.0%; Pred. No. 51;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 NLFETPIL 8
 DB 51 nlfetp11 58
 |||||
 |||stpl1 58
 RESULT 11
 AA044139
 ID AA044139 standard; protein; 306 AA.
 XX
 XX AA044139;
 AC
 XX 27-FEB-2002 (first entry)
 DT
 XX Propionibacterium acnes immunogenic protein #5035.
 DE
 XX SAPHO syndrome; synovitis, acne, pustulosis, hyperostosis, osteomyelitis;
 KW uveitis; endophthalmitis; bone, joint, central nervous system. Elisa.
 KW inflammatory lesion, acne vulgaris, enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 XX
 XX Propionibacterium acnes
 OS
 XX

PN W0200181581-A2.
 XX
 XX C1-NCV-2001.
 XX
 XX 20-APR-2001; 2001WO-0512865.
 XX
 XX 21-APR-2000; 2000US-194047P.
 PP
 PR 02-JUN-2000; 2000US-208841P.
 PR
 PR 07-JUL-2000; 2000US-216747P.
 XX
 XX (CORI-) CORIAX CORP.
 PA
 XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'Amisouneuve J, Zhang Y, Jen S, Carter D;
 XX
 XX WPI; 2001-616774/71.
 DR
 DR N-PSDB; AAS59521.
 XX
 XX Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris.
 XX
 XX Claim 3; SEQ ID No 5334; 1063pp; English.
 PS
 XX Sequences AAU9105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note, the sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX
 XX Sequence 306 AA;
 SQ
 Query Match 75.6%; Score 44; DB 22; Length 306;
 Best Local Similarity 66.7%; Pred. No. 82;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 NLFETPILA 9
 DB 40 nlfetp11 48
 |||||
 |||metp11 48
 RESULT 12
 AAB62694
 ID AAB62694 standard; peptide; 26 AA.
 XX
 XX AAB62694;
 AC
 XX 06-AUG-2001 (first entry)
 DT
 XX ABC1 protein external domain TM5 TM6 fragment (residues 795-820).
 DE
 XX ABC1, antilipemic, cholesterol, inhibitor, low density lipoprotein; LDL.
 KW Homo sapiens.
 OS
 XX W0200132184-A2.
 PN
 XX

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes that proper record-keeping is essential for transparency and accountability, particularly in financial matters. The text outlines various methods for organizing and storing records, including digital databases and physical filing systems. It also mentions the need for regular audits and reviews to ensure the integrity of the data.

2. The second part of the document focuses on the role of communication in achieving organizational goals. It highlights the importance of clear and concise communication, both internally and externally. The text provides guidelines for effective communication, such as using appropriate language, listening actively, and providing feedback. It also discusses the benefits of open communication, including improved collaboration and decision-making.

3. The third part of the document addresses the issue of risk management. It defines risk as the potential for loss or damage and explains how to identify, assess, and mitigate risks. The text provides a framework for risk management, including the identification of risks, the assessment of their likelihood and impact, and the implementation of control measures. It also discusses the importance of monitoring and reviewing risks over time.

4. The fourth part of the document discusses the importance of training and development. It emphasizes that ongoing training and development are essential for maintaining a skilled and motivated workforce. The text outlines various training methods, including classroom instruction, on-the-job training, and self-directed learning. It also discusses the importance of setting learning objectives and evaluating the effectiveness of training programs.

5. The fifth part of the document discusses the importance of innovation and creativity. It emphasizes that innovation and creativity are essential for staying competitive in a rapidly changing market. The text provides guidelines for fostering innovation and creativity, such as encouraging open-mindedness, providing resources, and creating a supportive environment. It also discusses the importance of protecting intellectual property and managing innovation projects.

6. The sixth part of the document discusses the importance of sustainability. It defines sustainability as the ability to meet the needs of the present without compromising the ability of future generations to meet their own needs. The text outlines various strategies for achieving sustainability, including reducing environmental impact, promoting social responsibility, and ensuring economic viability. It also discusses the importance of measuring and reporting on sustainability performance.

7. The seventh part of the document discusses the importance of ethics and governance. It emphasizes that ethics and governance are essential for building trust and credibility. The text provides guidelines for ethical behavior, such as being honest, transparent, and accountable. It also discusses the importance of establishing a strong governance structure and implementing effective controls.

8. The eighth part of the document discusses the importance of customer service. It emphasizes that excellent customer service is essential for retaining customers and attracting new ones. The text provides guidelines for providing excellent customer service, such as listening to customer feedback, resolving complaints quickly, and going above and beyond to meet customer needs. It also discusses the importance of training customer service staff and monitoring customer satisfaction.

9. The ninth part of the document discusses the importance of financial management. It emphasizes that sound financial management is essential for the long-term success of any organization. The text outlines various financial management practices, including budgeting, forecasting, and controlling. It also discusses the importance of maintaining accurate financial records and seeking professional advice when needed.

10. The tenth part of the document discusses the importance of strategic planning. It emphasizes that strategic planning is essential for setting clear goals and determining the best way to achieve them. The text provides guidelines for developing a strategic plan, such as conducting a SWOT analysis, setting SMART goals, and implementing the plan. It also discusses the importance of reviewing and updating the strategic plan regularly.

Query Match 80.0%; Score 36; DB 2; Length 612;
 Best Local Similarity 75.0%; Pred. No. 18;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPIL 8
 |||||
 Db 479 NLFETPIL 486

RESULT 3
 T48466
 hypothetical protein T1E3.60 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse ear cress)
 C:Date: 20 Apr 2000 #sequence_revision 20 Apr 2000 #text_change 20 Jul 2000
 C:Accession: T48466
 R:Bayan, M.; Terry, N.; Ardiles, W.; Buyssehart, C.; Passavillo, P.; Passavillo, P.; De
 cewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: 724491
 A:Accession: T48466
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-669 <REV>
 A:Cross-references: EMBL:AL162972
 A:Experimental source: cultivar Columbia; BAC clone T1E3
 C:Genetics:
 A:Map position: 5
 A:Introns: 7271-72671; 28741-30741; 32201; 46573
 A:Note: T1E3.60
 C:Superfamily: Arabidopsis thaliana hypothetical protein T1E3.60

Query Match 80.0%; Score 36; DB 2; Length 669;
 Best Local Similarity 75.0%; Pred. No. 20;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPIL 8
 |||||
 Db 112 NLFETPIL 119

RESULT 4
 T48464
 hypothetical protein T1E3.40 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse ear cress)
 C:Date: 20 Apr 2000 #sequence_revision 20 Apr 2000 #text_change 20 Apr 2000
 C:Accession: T48464
 R:Bayan, M.; Terry, N.; Ardiles, W.; Buyssehart, C.; Passavillo, P.; Passavillo, P.; De
 cewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: 724491
 A:Accession: T48464
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-705 <REV>
 A:Cross-references: EMBL:AL162972
 A:Experimental source: cultivar Columbia; BAC clone T1E3
 C:Genetics:
 A:Map position: 5
 A:Introns: 5971-28671; 28671-30741; 32201; 46573
 A:Note: T1E3.40

Query Match 80.0%; Score 36; DB 2; Length 705;
 Best Local Similarity 75.0%; Pred. No. 21;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPIL 8
 |||||
 Db 99 NLFETPIL 106

RESULT 5
 T11743
 pP47 protein - pig (fragment)
 C:Species: Sus scrofa domestica (domestic pig)
 C:Date: 16 Jul 1979 #sequence_revision 16 Jul 1999 #text_change 21-Jul-2000
 C:Accession: T11743
 R:Essling, M.; Vogel, T.; Calvete, J.J.; Thole, H.H.; Schmidtko, J.; Matsuda, T.; Foo
 Biol. Reprod. 58, 1057-1064, 1998
 A:Title: Molecular cloning and characterization of pP47, a novel boar sperm-associated
 A:Reference number: Z17325; NUID:98206817
 A:Accession: T11743
 A:Status: preliminary; translated from CP/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-409 <ENS>
 A:Cross-references: EMBL:Y11683; NID:12952927; FID:CAAG7079; FID:12952928
 A:Experimental source: testis
 C:Function:
 A:Description: may be involved in membrane remodeling and/or function as a zona pellu
 C:Superfamily: unassigned BCF-related proteins, BCF homology
 F.5.40/Domain: BCF homology - BCF

Query Match 77.8%; Score 45; DB 2; Length 409;
 Best Local Similarity 75.0%; Pred. No. 18;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLFETPIL 8
 |||||
 Db 378 NLFETPIL 385

RESULT 6
 T13799
 neurixin IV - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 13 Aug 1999 #sequence_revision 13 Aug 1999 #text_change 17 Nov 2000
 C:Accession: T13799
 R:Baumgartner, S.; Littleton, J.T.; Broadie, K.; Bhat, M.A.; Harbecke, K.; Lemay, J.
 Cell 87, 1059-1068, 1996
 A:Title: A Drosophila neurixin is required for septate junction and blood-nerve barrier
 A:Reference number: 217762; NUID:9714213
 A:Accession: T13799
 A:Status: preliminary; translated from GR/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1283 <BAU>
 A:Cross-references: EMBL:X86685; NID:31518220; FID:CAAG69841; FID:31518221
 C:Genetics:
 A:Gene: nrx
 A:Cross-references: FlyBase:FBgn014997
 C:Function:
 A:Description: is required for NRX in the formation of septate junction septa and int

Query Match 77.8%; Score 35; DB 2; Length 1283;
 Best Local Similarity 66.7%; Pred. No. 69;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLFETPIL 9
 |||||
 Db 153 NLFETPIL 161

RESULT 7
 S46082
 urea carboxylase (EC 6.3.4.1) yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein YBR1449; protein YBR208c; urea amidolyase
 C:Species: Saccharomyces cerevisiae
 C:Date: 26 Aug 1994 #sequence_revision 09 Sep 1994 #text_change 11 Jan 2002
 C:Accession: S46082; S46081; S46080; S41341; S34930; S4043
 R:Rieger, M.
 submitted to the Protein Sequence Database, August 1994
 A:Reference number: S45734
 A:Accession: S46082

A:Molecule type: DNA
 A:Residues: 1-893 <PE>
 A:Cross-references: EMBL:Z36077; MIPS:YBP208c
 A:Experimental source: strain S288c
 R:Feldmann, H.; Maunhaupt, G.; Schwarzslose, G.; Vetter, I.
 submitted to the Protein Sequence Database, August 1994
 A:Reference number: S45927
 A:Accession: S46081
 A:Molecule type: DNA
 A:Residues: 873-1835 <PE>
 A:Cross-references: EMBL:Z36077; MIPS:YBP208c
 A:Experimental source: strain S288c
 R:Bussereau, F.; Demolis, N.; Jacquet, M.; Mallet, L.
 submitted to the Protein Sequence Database, August 1994
 A:Reference number: S46054
 A:Accession: S46080
 A:Molecule type: DNA
 A:Residues: 1487-1835 <BOS>
 A:Cross-references: EMBL:Z36077; MIPS:YBP208c
 A:Experimental source: strain S288c
 R:Genbauffe, F.S.; Cooper, T.G.
 DNA Seq. 2; 14-32, 1991
 A:Title: The urea amidolyase (URE1.2) gene of *Saccharomyces cerevisiae*.
 A:Reference number: S31341; MIMD:92199240
 A:Accession: S41341
 A:Molecule type: DNA
 A:Residues: 1-95, 97-255, 257-258, 261-259, 458, 461-462, 829, 831-1394, 1417, 1396-1835
 A:Cross-references: EMBL:M44926; NID:J17312; PIDN:AC41643.1; PID:J73122
 R:Bussereau, F.; Mallet, L.; Gailliot, L.; Iriquet, M.
 Yeast 9; 797-806, 1993
 A:Title: Yeast Sequencing Reports. A 12.8 kb segment on the right arm of chromosome II
 A:Reference number: S34925; MIPS:Q3277417
 A:Accession: S34930
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1487-1835 <BOS>
 A:Cross-references: EMBL:Z21487; NID:J311675; PIDN:CAA7666.1; PID:J311784
 A:Experimental source: strain S288c
 C:Genetics:
 A:Geno: SGD:DUP1.2
 A:Cross-references: SGD:S0000412; MIPS:YBP208c
 A:Map position: 2F
 C:Superfamily: lipoyl/biotin-binding homology; biotin carboxylase homology
 C:Keywords: ATP, biotin binding, ligase, P loop, putative nucleotide binding; transmembran
 F:123-129/Region: nucleotide-binding motif A (P-loop)
 F:163-179/Domain: transmembrane #status predicted <TM>
 F:209-237/Domain: transmembrane #status predicted <TM>
 F:411-427/Domain: transmembrane #status predicted <TM>
 F:635-1082/Domain: biotin carboxylase homology P38
 F:1141-1157/Domain: transmembrane #status predicted <TM>
 F:1582-1598/Domain: transmembrane #status predicted <TM>
 F:1759-1832/Domain: lipoyl/biotin binding homology P46
 F:128/Binding site: ATP/GTP (lys) #status predicted
 F:1794/Binding site: biotin (lys) (covalent) #status predicted

Query Match 77.8%; Score 35; DB 2; Length 1835;
 Best Local Similarity 100.0%; Pred. No. 10; 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0

QY 3 FETPILA 9
 |||||
 DB 1418 FETPILA 1424

RESULT 8
 A12479
 C:Species: Anabaena sp.
 A:Note: Anabaena sp. (strain PCC 7120) is a synonym of *Nostoc* sp. strain PCC 7120
 C:Date: 14-Dec-2001; #sequence_revision: 14-Dec-2001; #text_change: 11-Jan-2002
 A:Accession: A12479
 R:Kanehisa, T.; Nakamura, Y.; Wolk, G.P.; Kurita, T.; Sasamoto, S.; Watanabe, A.; Iriquich

Rakaschi, R.; Shinto, S.; Sugimoto, M.; Takizawa, M.; Yamada, M.; Yasuda, M.; Tabata
 DNA Res. 2; 255-213, 2001
 A:Title: Complete Genome Sequence of the Filamentous Nitrogen Fixing Cyanobacterium
 A:Reference number: AB1807; MIMD:21595285; PMID:11759840
 A:Accession: A12479
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-142 <KE>
 A:Cross-references: GB:HA000020; PIDN:HAH78101.1; PID:q17135555; GSPDH:GN00180
 A:Experimental source: strain FCC 7120
 C:Genetics:
 A:Gene: alr7017
 A:Genome: plasmid

Query Match 75.6%; Score 34; DB 2; Length 142;
 Best Local Similarity 75.0%; Pred. No. 8;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MLETPIL 8
 |||||
 DB 65 MLETPIL 72

RESULT 9
 A70554
 A:Probable dehydrogenase Mycobacterium tuberculosis (strain H37PV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998; #sequence_revision: 17-Jul-1998; #text_change: 20-Sep-2000
 A:Accession: A70554
 R:Giles, S.L.; Busch, E.; Parkhill, J.; Gainer, T.; Charcher, C.; Harris, D.; Gordon
 Rajandream, M.A.; Koppes, J.; Rutter, S.; Seeger, K.; Skellern, S.; Squares, S.
 Nature 393; 537-544, 1998
 A:Authors: Squares, K.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Sequencing the Biology of Mycobacterium tuberculosis
 A:Reference number: A70554; MIMD:98295987
 A:Accession: A70554
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-250 <COL>
 A:Cross-references: GB:295584; GB:ALL23455; NID:J291774; PIDN:CA69932.1; PID:J21171
 A:Experimental source: strain H37PV
 C:Genetics:
 A:Gene: Kv1144
 C:Superfamily: ribitol dehydrogenase; short chain alcohol dehydrogenase homology
 F:1787/Domain: short chain alcohol dehydrogenase homology <SASH>

Query Match 75.6%; Score 24; DB 2; Length 250;
 Best Local Similarity 75.0%; Pred. No. 17;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LFTPTILA 9
 |||||
 DB 188 LFTPTILA 195

RESULT 10
 A12848
 A:ribose-phosphate pyrophosphokinase [imported] - Agrobacterium tumefaciens (strain C58
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002; #sequence_revision: 11-Jan-2002; #text_change: 31-Feb-2002
 A:Accession: A12848
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
 erick, G.; Gillet, W.; Grant, G.; Greenham, D.; Puyyappa, L.; Levy, R.; Li, M.; Moch
 Kaip, P.; Romero, P.; Zhang, S.
 Science 294; 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
 ster, R.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AR2577; PMID:11743193
 A:Accession: A12848

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-310 <GUA>
A:Cross-references: GB:AE000688; PIRN:AA143207.1; PDB:gl774688; GSFPB:GN00196
A:Experimental source: strain C58 (import)
C:Genetics:
A:Gene: prsA
A:Map position: circular chromosome
C:Superfamily: ribose-phosphate pyrophosphokinase catalytic chain

Query Match 75.6%; Score 34; DB 2; Length 310;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLEPPIIA 9
|||||
Db 141 NLEAAPPILA 149

RESULT 11
G97625
prsa (AF060605) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jun-2002
C:Accession: G97625
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Courjols, B.; Goldman, A.; Liu, P.; Wollam, C.; Allinger, M.; Doughty, P.; Scott, C.; Lappas, C.; Markelz, R.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium Tumefaciens
A:Reference number: A97359; PMID:11743194
A:Accession: G97625
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-318 <GUA>
A:Cross-references: GB:AE007869; PIRN:AAK87960.1; PDB:ql5157366; GSFPB:GN00169
C:Genetics:
A:Gene: AGR_C_4041
A:Map position: circular chromosome
C:Superfamily: ribose phosphate pyrophosphokinase catalytic chain

Query Match 75.6%; Score 34; DB 2; Length 318;
Best Local Similarity 77.9%; Pred. No. 22;
Matches 7; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLEPPIIA 9
|||||
Db 149 NLEAAPPILA 157

RESULT 12
A11403
dihydroorotate dehydrogenase homolog pyrd [imported] - Listeria monocytogenes (strain EG)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27 Nov 2001 #text_change 14-Dec-2001
C:Accession: A11403
R:Glaser, P.; Franquell, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, F.; Bleecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Bussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.; Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Madueno, E.; Maitournam, A.; Mok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla, A.; Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: A11403
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-304 <GUA>
A:Cross-references: GB:NC_003210; PIRN:CA099911.1; PDB:ql6411287; GSFPB:GN00177
A:Experimental source: strain EG0-e
C:Genetics:
A:Gene: pyrd

C:Superfamily: dihydroorotate oxidase

Query Match 73.4%; Score 33; DB 2; Length 304;
Best Local Similarity 85.7%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 FETPIIA 9
|||||
Db 92 FETPIIA 98

RESULT 13
A11675
dihydroorotate dehydrogenase homolog pyrd [imported] - Listeria innocua (strain Cliph)
C:Species: Listeria innocua
C:Date: 27 Nov 2001 #sequence_revision 27 Nov 2001 #text_change 14-Dec-2001
C:Accession: A11675
R:Glaser, P.; Franquell, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, F.; Bleecker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Bussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.; Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maitournam, A.; Mok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla, A.; Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: A11675
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-304 <GUA>
A:Cross-references: GB:AL592022; PIRN:CA097177.1; PDB:ql6411449; GSFPB:GN00198
A:Experimental source: strain Cliph1262
C:Genetics:
A:Gene: pyrd
C:Superfamily: dihydroorotate oxidase

Query Match 73.4%; Score 33; DB 2; Length 304;
Best Local Similarity 85.7%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 FETPIIA 9
|||||
Db 92 FETPIIA 98

RESULT 14
T37716
actin-interacting protein bud6/Aip3p protein SPAC15E1.01 [imported] - fission yeast (C:Species: Schizosaccharomyces pombe
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
C:Accession: T37716
R:Murphy, L.; Harris, D.; Barrell, H.G.; Rajandream, M.A.; Lyne, M.H.; submitted to the EMBL Data Library, August 1999
A:Reference number: Z21739
A:Accession: T37716
A:Status: preliminary; translated from GS/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-363 <MUR>
A:Cross-references: EMBL:AL05770; PIRN:CA052420.1; GSFPB:GN00066; SPDB:SPAC15E1.01
A:Experimental source: strain 972h; cosmid cl5E1
C:Genetics:
A:Gene: SPAC15A.0.15, SPDB:SPAC15E1.01
A:Map position: 1

Query Match 73.3%; Score 33; DB 2; Length 363;
Best Local Similarity 82.5%; Pred. No. 42;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLEPPIIA 8
|||||
Db 282 NLEPPIIV 289

Thu Sep 5 16:28:47 2002

RESULT 15

S65138 glycoprotein antigen MGP57/53, mammary gland - bovine (fragment)
 N:Alternate names: glycoprotein component 16/major fat-globule membrane protein/MRG-EB8
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 28-Oct-1996 #sequence_revision 13 Mar 1997 #text_change 07-Aug-1998
 C:Accession: S65138; G48394
 R:Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, E.; Matsuda, T.
 Biochim. Biophys. Acta 1245: 385-391, 1995
 A:Title: Molecular cloning of glycoprotein antigens MGP57/53 recognized by monoclonal antibodies
 A:Reference number: S65138; MUID:96125736
 A:Accession: S65138
 A>Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-401 <AOK>
 R:Mathar, I.H.; Banhart, L.R.; Lane, W.S.
 Biochem. Mol. Biol. Int. 29: 545-554, 1993
 A:Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea pig
 I1-like sequences.
 A:Reference number: A48394; MUID:93280576
 A:Accession: G48394
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 207-220 <MAT>
 A:Experimental source: milk
 A:Note: sequence extracted from NCBI backbone (NCBI:P131457)
 C:Superfamily: milk fat globule protein, discoidin I amino-terminal homology, EGF homolo
 C:Keywords: glycoprotein
 F:1-35/Domain: EGF homology (fragment) <EGL
 F:40-76/Domain: EGF homology <EG2
 F:82-229/Domain: discoidin I amino-terminal homology <DRI
 F:243-401/Domain: discoidin I amino-terminal homology <DN2>

Query Match 73.3% Score 33; DB 2; Length 401;
 Best Local Similarity 66.7%; Pred. No. 48;
 Matches 6; Conservative 1; Mismatches 2; Gaps 0;

QY 1 NLFETPILIA 9
 |:|||||
 Db 370 NIFETPFOA 378

Search completed: September 5, 2002, 15:28:51
 Job time: 356 sec



GenCore version 4.5
Copyright (c) 1994 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run ID: 547000015, 2602, 15, 31, 40, Search time 43.65 seconds
(without alignments)
10.252 Million cell updates/sec

Title: US-09-744-804-37
Perfect score: 45
Sequence: 1 NLEPILPIL 9

Scoring table: BLOSUM62

Gapop 10 0, Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	45	100.0	387	1	Q08431 homo sapien
2	36	80.0	612	1	P55782 gadus morhu
3	35	77.8	409	1	P79385 sus scrofa
4	35	77.8	1284	1	Q64887 drosophila
5	35	77.8	1835	1	D011_YEAST
6	33	73.3	427	1	Q95114 bos taurus
7	33	73.3	464	1	Q10311 canorhabdi
8	33	73.3	535	1	SPRC_SYNT3
9	33	73.3	1385	1	FAT1_SCHPO
10	33	73.3	1416	1	RLM_MOUSE
11	32	71.1	527	1	YAD7_SCHFG
12	32	71.1	532	1	Q09033 schistosom
13	32	71.1	553	1	Q09033 schistosom
14	32	71.1	2261	1	Q09033 schistosom
15	32	71.1	2261	1	Q09033 schistosom
16	32	71.1	3801	1	Q09033 schistosom
17	31	68.9	98	1	Y528_METJA
18	31	68.9	255	1	HCD2_DROME
19	31	68.9	333	1	Q10311 canorhabdi
20	31	68.9	366	1	Q09033 schistosom
21	31	68.9	427	1	Q09033 schistosom
22	31	68.9	463	1	Q09033 schistosom
23	31	68.9	665	1	Q09033 schistosom
24	31	68.9	1062	1	Q09033 schistosom
25	30	66.7	81	1	Q09033 schistosom
26	30	66.7	81	1	Q09033 schistosom
27	30	66.7	195	1	Q09033 schistosom
28	30	66.7	265	1	Q09033 schistosom
29	30	66.7	311	1	Q09033 schistosom
30	30	66.7	322	1	Q09033 schistosom
31	30	66.7	489	1	Q09033 schistosom
32	30	66.7	395	1	Q09033 schistosom
33	30	66.7	418	1	Q09033 schistosom

34	30	66.7	441	1	GAT6_RAT
35	30	66.7	444	1	GAT6_MOUSE
36	30	66.7	449	1	GAT6_HUMAN
37	30	66.7	485	1	V136_IRBPA
38	30	66.7	492	1	GATB_ARCFU
39	30	66.7	504	1	V104_MYCTH
40	30	66.7	516	1	Y1J7_YEAST
41	30	66.7	540	1	V112_VANCP
42	30	66.7	546	1	V112_VANCP
43	30	66.7	535	1	V812_VAFV
44	30	66.7	652	1	YD84_HUMAN
45	30	66.7	703	1	PALY_BROFI

ALIGNMENTS

RESULT 1					
MFGM_HUMAN					
AC Q08431					
DT 01-OCT-1996 (rel. 34, Created)					
DT 01 NOV 1997 (rel. 35, Last sequence update)					
DE 16-OCT-2001 (rel. 40, Last annotation update)					
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (HMFG)					
DE (Breast epithelial antigen BA46) (MFGM) [Contains: Medin].					
GN MFGF8.					
OS Homo sapiens (Human).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX NCBI_taxID=9606;					
RN [1]					
RP SEQUENCE FROM N.A.					
RC MEDLINE=6621406, PubMed=8609264;					
RA Gotoh T.P., Taylor M.P., Godwin S.G., Ceriani P.L., Peterson J.A.;					
RT "Cloning and sequence analysis of human breast epithelial antigen					
RT BA46 reveals an RGD cell adhesion sequence presented on an epidermal					
RT growth factor-like domain";					
RL DNA Cell Biol. 15:281-286(1996).					
RN [2]					
RP SEQUENCE OF 170-487 FROM N.A.					
RC TISSUE=Mammary gland;					
RA Gotoh T.P., Taylor M.P., Godwin S.G., Ceriani P.L., Peterson J.A.;					
RT "Cloning and sequence analysis of human breast epithelial antigen					
RT BA46 reveals an RGD cell adhesion sequence presented on an epidermal					
RT growth factor-like domain";					
RL DNA Cell Biol. 15:281-286(1996).					
RN [3]					
RP PARTIAL SEQUENCE, AND CHARACTERIZATION.					
RC TISSUE=Milk;					
RA Gotoh T.P., Taylor M.P., Godwin S.G., Ceriani P.L., Peterson J.A.;					
RT "Cloning and sequence analysis of human breast epithelial antigen					
RT BA46 reveals an RGD cell adhesion sequence presented on an epidermal					
RT growth factor-like domain";					
RL DNA Cell Biol. 15:281-286(1996).					
RN [4]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [5]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [6]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [7]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [8]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [9]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [10]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [11]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [12]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [13]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [14]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [15]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [16]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [17]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [18]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [19]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [20]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [21]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [22]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [23]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [24]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					
RT breast carcinoma protein BA46 from human milk fat globule membranes";					
RL J. Protein Chem. 17:143-148(1998).					
RN [25]					
RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.					
RC MEDLINE=9444507, PubMed=10411933;					
RA Haegele J.E., Naeffels J., Sletten E., Westermark G., Westermark P.;					
RT "Isolation and characterization of full and truncated forms of human					

BL DNA Cell Biol 16:861-869(1997).
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
 CC SPECIFICALLY TO 6-ANILINOS AND INHERITS ITS REPLICATION.
 CC -1- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIA. AMYLOID
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
 CC MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
 CC -1- PTM: MEDIN HAS A RACGED N-TERMINUS WITH MINOR SPECIES STARTING AT
 CC AMINO ACID 264 AND 273.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 P5/8 TYPE C DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/autumn97/>
 CC or send an email to license@sib-sib.ch).
 DR EMBL: 058516; AA:58549; 1; -
 DR EMBL: 056151; AAR19771; 1; -
 DR MIM: 657283;
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000421; P58_C.
 DR Pfam: PF00008; EGF; 1.
 DR Pfam: PF00754; P5/P8_Type_C; 2
 DR SMART: SM00181; EGF_1
 DR SMART: SM00231; P58C; 2.
 DR PROSITE: PS00077; EGF_1; 1
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01285; P58C; 1; 2.
 DR PROSITE: PS01286; P58C; 2; 2
 KW Signal: Glycoprotein, MLT, repeat, EGF-like domain, Amyloid
 FT SIGNAL 1 23
 FT CHAIN 24 387 LACTADHERIN.
 FT CHAIN 295 387 LACTADHERIN; SHORT FORM
 FT CHAIN 268 317 MEDIN.
 FT DOMAIN 24 67 EGF-LIKE.
 FT DOMAIN 70 225 F5/8 TYPE C 1.
 FT DOMAIN 230 387 F5/8 TYPE C 2.
 FT SITE 46 48 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 27 38 BY SIMILARITY.
 FT DISULFID 32 55 BY SIMILARITY.
 FT DISULFID 57 66 BY SIMILARITY.
 FT DISULFID 70 225 BY SIMILARITY.
 FT DISULFID 212 216 BY SIMILARITY.
 FT DISULFID 230 387 BY SIMILARITY.
 FT CARBOHYD 238 238 N-LINKED (GLUCNA...) (POTENTIAL).
 FT CARBOHYD 325 325 N-LINKED (GLUCNA...) (POTENTIAL).
 FT CARBOHYD 329 329 N-LINKED (GLUCNA...) (POTENTIAL).
 FT CARBOHYD 350 350 N-LINKED (GLUCNA...) (POTENTIAL).
 SQ SEQUENCE 387 AA, 43123 MW, 2895571 Da; 637820 Cys64.
 Query Match 100.0%; Score 45; DR 1; Length 387;
 Best Local Similarity 100.0%; Pred No. 6.0;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NLFETPILA 9
 I|||||
 DB 356 NLFETPILA 364
 RESULT 2
 ID NUFM_GADMO STANDARD; PPT; 612 AA.
 AC P55782;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE NADH-ubiquinone oxidoreductase chain 5 (EC 1.6.5.3)
 GN MTND5 OR ND5.
 This SWISS-PROT entry is copyright. It is produced through a collaboration

Gadus morhua (Atlantic cod).
 CC Mitochondrion.
 CC Fokaryota, Metazoa, Chordata, Craniata, Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 CC Acanthopterygii; Paracanthopterygii; Gadiformes; Gadidae; Gadidae;
 CC Gadus.
 CC NCBI_TaxID=8049;
 RN [1]
 PP SEQUENCE FROM N.A.
 EC SPAIN-NORWEGIAN COASTAL 1;
 EC MEDLINE=4641425; PubMed=2817426;
 PA Johansen S., Rakke I.;
 FT "The complete mitochondrial DNA sequence of Atlantic cod (Gadus
 FT morhua): relevance to taxonomic studies among codfishes.";
 FT Mol Mar Biol Biotechnol. 5:203-214(1996).
 CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone -> NAD(+) + ubiquinol.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/autumn97/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL: X99772; CAA68115; 1;
 DR InterPro: IPR003916; NADH_oxidore5.
 DR InterPro: IPR001750; Oxidored_q1.
 DR InterPro: IPR001516; Oxidored_q1_N
 DR Pfam: PF00361; Oxidored_q1; 1.
 DR Pfam: PF00662; Oxidored_q1_N; 1.
 DR PRINTS: PR01434; NADHGNASE5.
 KW Oxidoreductase; NAD; ubiquinone; Mitochondrion; Transmembrane.
 SQ SEQUENCE 612 AA, 66543 MW, 44694 Da; 15150 Cys64;
 Query Match 100.0%; Score 45; DR 1; Length 612;
 Best Local Similarity 75.0%; Pred No. 7.7;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NLFETPIL 8
 I|||||
 DB 479 NLFETPIL 486
 RESULT 3
 ID MFCM_PIG STANDARD; PPT; 409 AA
 AC P79385;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-MAY-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Lactadherin (Milk fat globule-EGF factor 8) (MFG-E8) (MFCM) (Sperm
 DE surface protein SP47) (PP47).
 DE MFG8.
 GN Sus scrofa (pig).
 CC Fokaryota, Metazoa, Chordata, Craniata, Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus
 CC NCBI_TaxID=9823;
 RN [1]
 PP SEQUENCE FROM N.A.
 EC TISSUE=Testis;
 EC Euslin M.;
 FT Submitted (NCV-1997) to the EMBL/GenBank/DBJ databases.
 FT -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
 FT BINDING PROTEIN.
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
 CC SPERMATOZOAN.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 P5/8 TYPE C DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, repackaged, or for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/press/pr000001>)
 CC or send an email to license@ebi.ac.uk).

CC EMBL: Y11683; CAA72374.1; -
 DR HSSP: P00740; IEDM
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000421; FASP.
 DR Pfam: PF00008; EGF; 2.
 DR Pfam: PF00754; F5_F8_type_C; 2.
 DR SMART: SM00181; EGF; 2.
 DR SMART: SM00231; FAS8C; 2.
 DR PROSITE: PS00022; EGF_1; 2.
 DR PROSITE: PS01184; EGF_2; 2.
 DR PROSITE: PS01295; FAS8C_1; 2.
 DR PROSITE: PS01296; FAS8C_2; 2.
 KW Glycoprotein; Repeat; EGF type domain.
 FT DOMAIN 2 41 EGF-like 1.
 FT DOMAIN 44 88 EGF-like 2.
 FT DOMAIN 91 247 F5/F8 type C 1.
 FT DOMAIN 252 409 F5/F8 type C 2.
 FT SITE 67 64 CELL ATTACHMENT SITE (POTENTIAL)
 FT DISULFID 6 17 BY SIMILARITY.
 FT DISULFID 11 29 BY SIMILARITY.
 FT DISULFID 31 40 BY SIMILARITY.
 FT DISULFID 91 247 BY SIMILARITY.
 FT DISULFID 244 238 BY SIMILARITY.
 FT DISULFID 252 409 BY SIMILARITY.
 FT CARBOHYD 41 41 N-LINKED (GLCNAG...) (POTENTIAL).
 FT CARBOHYD 372 372 N-LINKED (GLCNAG...) (POTENTIAL).
 SQ SEQUENCE 409 AA: 45725 MW: 80674P9039927A CP64.

Query Match 77.88; Score 35; HPI 1; Length 409;

Best Local Similarity 75.04; Pred. No. 8.2;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0.

QY 1 NIFETPIL 8
 I:|||||
 DB 378 NMFETPIL 365

RESULT 4
 ID NRX4_DROME STANDARD; PRT: 1284 AA.
 AC Q94887; G9VTU5;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Neurexin IV precursor
 GN NRX OR CG6827.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta.
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN Canton-S.
 RX MEDLINE=9713213; PubMed 8978610;
 RA Baumgartner S.W., Littleton J.T., Broadie K., Bhat M.A., Harbecke R.,
 RA Lenquet J.A., Chiquet-Ehrismann R., Prokop A., Rellen H.J.;
 RT "A Drosophila neurexin is required for septate junction and blood-
 RT nerve barrier formation and function".
 RL Cell 87:1059-1068(1996).
 PN [2]
 PP SEQUENCE FROM N.A
 RC STEADY-STATE
 RX MEDLINE=20196006; PubMed-10731132;
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gorayne J.D.,

KA Amanatides P.G., Scherer S.E., Li P.W., Hoskins K.A., Calle K.F.,
 KA George K.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
 KA Sutton G.G., Wirtzen J.R., Vandell M.D., Zhang Q., Chen L.X.,
 KA Branden R.G., Kisters V.H.C., Blazek R.G., Champe M., Pfeiffer E.D.,
 KA W. P. H. L., P. H. L., H. L., H. L., H. L., H. L., H. L., H. L.,
 KA April J.F., Aikayani A., An H.-J., Andrews-Pfaukoch C., Haldwin D.,
 KA Ballow P.M., Basu A., Bayandale J., Bayraktaroglu L., Bosley E.M.,
 KA Beeson K.Y., Beus P.V., Norman R.P., Bhandari D., Polshakov S.,
 KA Borkova D., Bouchan M.P., Roush J., Rokstein P., Brothier P.,
 KA Butts K.C., Busam D.A., Butler H., Cadieu E., Genter A., Chandra L.,
 KA Chert, C.M., Cady, S., Lahtke C., Gachet L.B., Gales P.,
 KA de Pablos B., Belcher A., Beng Z., Mays A.B., Hew L., Dietz S.M.,
 KA Dodson E., Doolittle R., Edwards M., Eppan, E.A., Eppan, E.A., Eppan, E.A.,
 KA Gordin K.J., Gnanapavan S., Fetzl C., Fetzl C., Fetzl C., Fetzl C.,
 KA Glesler C., Gabriellian A.E., Garg N.S., Gelhart W.M., Glasner K.,
 KA Godek A., Gao F., Gorielli J.H., Gu Z., Guan P., Harris M.,
 KA Harris R.L., Harvey D., Heiman T., Hernandez J.K., Heek J.,
 KA Hostin D., Houston K.A., Howard T., Wei M.H., Hedgum C.,
 KA Jafari M., Kalush E., Kaipen G.H., K. F. 7, Kornelson J.A., Korchum K.A.,
 KA Kama, P.L., Kadisa G., Kraft G., Kraybill S., Kraybill S., Kraybill S.,
 KA Laska F., Lei Y., Leistikow A.A., Li J., Li J., Li J., Li J.,
 KA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 KA Merkulov G., Milshina N.V., Mohanty C., Morris J., Moshrefi A.,
 KA Mout R.M., M. Y. M., Murphy K., Murphy K., Murphy K., Murphy K.,
 KA Nelson E.F., Nelson K.A., Nish K., Russell E.F., Paele J.M.,
 KA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 KA Pelner K., Remington K., Saunders P.D.C., Scheeler F., Shen H.,
 KA Shue B.C., Sidun-Klamos I., Simpson M., Skupski M.P., Smith T.,
 KA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 KA Swirskas F., Teller C., Turner P., Venter E., Wang A.H., Wang X.,
 KA Wang Z. Y., Wassenaar D.A., Weinstein G.M., Weisselbach J.,
 KA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 KA Ye X., Yeh F., Yeh F., Yeh F., Yeh F., Yeh F., Yeh F., Yeh F.,
 KA Zhang X.H., Zhang F.N., Zhang W., Zhang X., Zhang X., Zhang X.,
 KA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [3]
 RN INTERACTION WITH CORACLE.
 RX MEDLINE 9817173; PubMed 9508778;
 RA Ward R.E. IV, Lamb R.S., Fehon R.G.;
 RT "A conserved functional domain of Drosophila coracle is required for
 RT localization at the septate junction and has membrane-organizing
 RT activity.";
 RL J. Cell Biol. 140:1463-1473(1998).
 RN [4]
 RN INTERACTION WITH DISCS LOST.
 RX MEDLINE 9925024; PubMed 10302271;
 KA Bhat M.A., Izaddoust S., Lu Y., Cho K.O., Choi K.W., Rellen H.J.;
 RT "Discs lost, a novel multi PLE domain protein, establishes and
 RT maintains epithelial polarity.";
 RL Cell 96:833-845(1999).
 CC -1- FUNCTION: SEEMS TO PLAY A ROLE IN THE FORMATION AND FUNCTION OF
 CC SEPTATE JUNCTIONS. IS REQUIRED FOR THE BLOOD-BRAIN BARRIER
 CC FORMATION.
 CC -1- SUBUNIT: THE C-TERMINAL REGION INTERACTS WITH CORACLE. INTERACTS
 CC WITH DISCS LOST IN CIS FORM.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein (potential).
 CC -1- TISSUE SPECIFICITY: FOUND IN SEPTATE JUNCTIONS OF EPITHELIAL AND
 CC GLIAL CELLS.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 F5/F8 TYPE C DOMAIN.
 CC -1- SIMILARITY: CONTAINS 4 LAMININ G-LIKE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE NEUREXIN FAMILY.
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, repackaged, or for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/press/pr000001>)
 CC or send an email to license@ebi.ac.uk).

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U29412; AAC44566.1;
 DR WormBase: T19C3.5; CE02056;
 DR InterPro: IPR001124; LRP_BPL_CETP.
 DR Pfam: PF01273; LRP_BPL_CETP; 1.
 DR Pfam: PF02886; LRP_BPL_CETP; 1.
 DR SMART: SM00328; BP11; 1.
 DR SMART: SM00329; BP12; 1.
 DR PROSITE: PS00400; LRP_BPL_CETP; FALSE_NEG.
 KW Hypothetical protein; Signal.
 FT SIGNAL 1 17 PESTITIAL
 FT CHAIN 18 464 HYPOHETICAL PROTEIN T19C3.5.
 SQ SEQUENCE 464 AA: 51939 MW: 4029CF072F5E67F CRC64;

Query Match 73.3%; Score 33; DB 1; Length 464;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETP 6
 |||||
 DB 134 NLFETP 139

RESULT 8
 SPKC_SYNY3 STAN:ABD; PPT; 535 AA.
 AC P74745;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable serine/threonine-protein kinase C (EC 2.7.11.-).
 GN SPKC OR SIK0599.
 OS Synochocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synochocystis.
 OX NCBI_TaxID:1148.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kamei A., Ikouchi M.;
 PT "A novel gene, *SPKC*, encodes active Ser/Thr protein kinase in the
 PT motile cyanobacterium *Synochocystis* sp. PCC 6803."
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Kaneko T., Sato S., Kojima H., Tanaka A., Asamiz E., Nakamura Y.,
 RA Miyajima N., Hirosewa M., Sudlira M., Sasamoto S., Kimura T.,
 PA Inoshima T., Maruyama A., Murayama A., Nakagaki N., Nakano K., Okumura S.,
 RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT *Synochocystis* sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions."
 RL DNA Res. 3:109-136(1996).
 CC -1- SIMILARITY: BELONGS TO THE SPK/THR FAMILY OF PROTEIN KINASES.
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AB046599; BAB17035.1;
 SQ SEQUENCE 1385 AA: 154325 MW: AB83D40C4FE7537 CRC64;

DR EMBL: D90917; BAA18865.1;
 CE InterPro: IPR005719; Euk_pkinase.
 DE InterPro: IPR005719; Ser_thr_kinase.
 DE Pfam: PF00069; pkinase; 1.
 DE PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
 DE PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DE PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 KW Transferase; Serine/threonine-protein kinase; ATP-binding;
 KW Complete proteome.
 FT DOMAIN 12 277 PROTEIN KINASE.
 FT NP_BIND 18 26 ATP (BY SIMILARITY).
 FT BINDING 43 43 ATP (BY SIMILARITY).
 FT ACT_SITE 142 142 BY SIMILARITY.
 SQ SEQUENCE 535 AA: 58141 MW: 533C9F0E0E4D24D8 CRC64;

Query Match 73.3%; Score 33; DB 1; Length 535;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETP 6
 |||||
 DB 396 NLFETP 401

RESULT 9
 FATL_SCHPO STANDARD; PPT; 1485 AA.
 ID FATL_SCHPO
 AC O13735; G90UT0;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Actin interacting protein 3 homolog.
 GN FAT1 OR SPAC15A15 IS OR SPAN15P1.01.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OX NCBI_TaxID:4896;
 RN [1]
 RP SEQUENCE OF 1-1033 FROM N.A.
 RC STEAIN-972;
 RA Murphy L., Harris D., Wood V., Barrell B.G., Rajandream M.A.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1023-1385 FROM N.A.
 RC STRAIN-972;
 RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Lyne M.H.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP IDENTIFICATION, AND GENE NAME.
 EX MEDLINE-20143685; PubMed-10679021;
 RA Jin H., Amberg D.C.;
 RT "The secretory pathway mediates localization of the cell polarity
 RT regulator *Apip/Rud6p*."
 FT Mol Biol Cell 11:147-161(2000).
 CC -1- FUNCTION: INVOLVED IN THE ORGANIZATION AND/OR FUNCTION OF THE
 CC ACTIN CYTOSKELETON.
 CC -1- SIMILARITY: TO YPAST RUD6
 CC
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: Z97208; CAB13112.1;
 DE A10047; A10047.1;
 KW Coiled coil; Cytoskeleton.
 FT DOMAIN 1009 1096
 SQ SEQUENCE 1385 AA: 154325 MW: AB83D40C4FE7537 CRC64;

DR PROSITE: PS01241; TRMA-2; 1.
 KW Hypothetical protein; Hydrolase; Nuclease; Transferase;
 KW Metalloproteinase 479 BY SIMILARITY.
 FT ACT_SITE 479
 SQ SEQUENCE 527 AA; 59613 MW; ACC640D6C6FFK77F CRC64;

Query Match 71.1%; Score 32; DB 1; Length 527;
 Best Local Similarity 66.7%; Pred. No. 45;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLFETPILA 9
 ||| 1 1 1 1
 Db 96 NLFETPILA 104

RESULT 12
 ICP0_HSVB STANDARD; PRT; 532 AA;
 AC P28940;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Trans-acting transcriptional protein ICP0.
 GN 63.
 OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OF Alphaherpesvirinae; Varicellovirus.
 OX NCBI_TaxID 31520.
 RN [1]
 RP MEDLINE 92295565; PubMed 1181806;
 RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
 RT "The DNA sequence of equine herpesvirus-1.";
 RL Virology 189:304-316(1992).
 RN [2]
 RP STRUCTURE BY NMR OF 1-63
 RX MEDLINE 94087718; PubMed 8263911;
 RA Everett R.D., Barlow P.N., Milner A., Luisi B., Orr A., Hope G.,
 RA Lyne P.;
 RT "A novel arrangement of zinc binding residues and secondary structure
 in the CHC4 motif of an alpha herpes virus protein family.";
 RL J. Mol. Biol. 234:1038-1047(1994).
 RN [3]
 RP STRUCTURE BY NMR OF 1-63.
 RX MEDLINE 94172642; PubMed 8126734;
 RA Barlow P.N., Luisi B., Milner A., Elliott M., Everett R.D.;
 RT "Structure of the CHC4 domain by 1H nuclear magnetic resonance
 spectroscopy. A new structural class of zinc-finger.";
 PL J. Mol. Biol. 247:261-271(1994).
 CC 1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 CC 1- SIMILARITY: TO OTHER HERPESVIRUSES ICP0 PROTEIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed, usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/atomnew/>
 or send an email to license@isb-sib.ch).

EMBL: M86664; AAR02498 1;
 FIR: 136801; W2HEF5.
 DR PDB: 1CHC; 30-APR-94.
 DR InterPro: IPR001841; Znf_ring.
 DR Pfam: PF00097; zf-C3HC4; 1.
 DR SMART: SM00184; KING; 1.
 DR PROSITE: PS00518; ZF_RING_1; 1.
 DR PROSITE: PS00699; ZF_RING_2; 1.
 FW Transcription regulation, Trans acting factor, Activator, Zinc finger
 DNA-binding; 4D-structure.
 KW ZN_FING 8 47 RING-TYPE.
 FT METAL 8 8 ZINC 1.

FT METAL 11 11 ZINC 1.
 FT METAL 24 24 ZINC 2.
 FT METAL 26 26 ZINC 2.
 FT METAL 29 29 ZINC 1.
 FT METAL 32 32 ZINC 1.
 FT METAL 43 43 ZINC 2.
 FT METAL 46 46 ZINC 2.
 FT LORAIN 219 217 POLY-SER.
 FT STRAND 19 21
 FT TURN 22 25
 FT STRAND 26 28
 FT TURN 30 31
 FT HELIX 32 39
 FT TURN 43 46
 FT STRAND 53 54
 SQ SEQUENCE 532 AA; 58629 MW; B4CB7E16FA26FDFA CRC64;

Query Match 71.1%; Score 32; DB 1; Length 532;
 Best Local Similarity 87.5%; Pred. No. 46;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFETPILA 9
 ||||| 1 1
 Db 136 LFETPILA 143

RESULT 13
 ODO2_MYCTU STANDARD; PRT; 553 AA.
 AC Q10381;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Dihydrolipoamide succinyltransferase component of 2-oxoglutarate
 DE dehydrogenase complex (EC 2.3.1.61) (E2).
 GN SUCR OF PW2215 OF MT2272 OF MTCV190.26.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID-1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE-98295987; PubMed-9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Radoock K., Hasham D., Brown D., Chillingworth T., Connor R.,
 RA Davies K., Devlin K., Pettell T., Gentles S., Hamlin N., Helgoyd S.,
 RA Hensby T., Jagels K., Kregh A., McLean J., Meule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Ralandream M.A., Rogers J.,
 RA Rutter S., Senger K., Skellton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;

"Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.

RA STRAIN-H37RV;
 RX MEDLINE-1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Fumolacca M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains.";
 RL Submitted (APR-2001) to the FMBI/GenBank/EMBL databases.

CC 1- FUNCTION: THE 2-OXOGLUTARATE DEHYDROGENASE COMPLEX CATALYZES THE
 OVERALL CONVERSION OF 2-OXOGLUTARATE TO SUCCINYL-CoA & CO(2). IT
 CONTAINS MULTIPLE COPIES OF 3 ENZYMIC COMPONENTS: 2-OXOGLUTARATE
 DEHYDROGENASE (E1), DIHYDROLIPOAMIDE SUCCINYLTRANSFERASE (E2) AND
 LIPOLAMIDE DEHYDROGENASE (E3) (BY SIMILARITY).
 CC 1- CATALYTIC ACTIVITY: Succinyl-CoA + dihydrolipoamide - CoA + S-

RX MEDLINE 20549092; PubMed 11086027;
 RA Clee S.M., Kastelein J.J.P., van Dam M., Marcil M., Poonp K.,
 RA Zwarts K.Y., Collins J.A., Beekmans R., Tanusawa N., Suijs T.,
 RA Suda T., Gaska R., Boucher B., Pondeau C., Desouch C.,
 RA Brooks Wilson A., Mellisham H.C.F., Fehlich J., Ganes J., Jr.,
 RA Hayden M.K.;
 RT "Age and residual cholesterol efflux affect HDL cholesterol levels and
 RT coronary artery disease in ABCA1 heterozygotes";
 RL J. Clin. Invest. 106:1263-1276(2000)
 RN [11]
 RP VARIANT ID ASN-1289 AND HIS-1800;
 RX MEDLINE 21171544; PubMed 10795561;
 RA Brousseau M.E., Schaefer E.J., Dupuis J., Eustace B.,
 RA Van Berwenh P., Goldkamp A.L., Thurston L.M., Fitzgerald M.G.,
 RA Yask McKenna D., O'Reilly G., Elfarf G.F., Wellenbach B.,
 RA Ordozav J.M., Freeman M.W., Brown R.H., Jr., Gu J.Z.;
 RA "Novel mutations in the gene encoding ATP binding cassette 1 in four
 RT disorder kindreds";
 RL J. Lipid Res. 41:433-441(2000)
 RN [12]
 RP VARIANT ID ASP-1046, VARIANT IDA CYS-200, AND VARIANTS IYS-219,
 RP ILE-825; MET-883 AND IYS-1587;
 RX MEDLINE-zu06033; PubMed-10946021;
 RA Wang J., Burnett J.R., Near S., Young K., Zimman B., Hanley A.J.G.,
 RA Connolly P.W., Harris S.B., Hegele R.A.;
 RA "Common and rare ABCA1 variants affecting plasma HDL cholesterol";
 RL Arterioscler. Thromb. Vasc. Biol. 20:1044-1049(2000)
 RN [13]
 RP VARIANT ID TRP-587, AND VARIANT LEU-2168;
 RX MEDLINE-21157002; PubMed 11257260;
 RA Bertolini S., Pisciotto L., Seri M., Cusano K., Cantafora A.,
 RA Calabresi L., Franceschini G., Parascenzo R., Calandra S.;
 RA "A point mutation in ABCA1 gene in a patient with severe premature
 RT coronary heart disease and mild clinical phenotype of Tangier
 RT disease";
 RL Atherosclerosis 154:599-605(2001).
 RN [14]
 RP VARIANTS IYS-219; MET-983 AND ASP-1172;
 RX MEDLINE 21157003; PubMed 11257261;
 RA Brousseau M.E., Rodziach M., Schaefer E.J., Goldkamp A.L., Kielar D.,
 RA Probst M., Ordozav J.M., Aslanidis C., Lackner K.J.,
 RA Bloomfield Rubins H., Collins D., Robins S.J., Wilson P.W.F.,
 RA Schmitz G.;
 RA "Common variants in the gene encoding ABC binding cassette transporter
 RT 1 in men with low HDL cholesterol levels and coronary heart disease";
 RL Atherosclerosis 154:607-611(2001).
 RN [15]
 RP VARIANT ID LEU-1506;
 RX MEDLINE 21369429; PubMed 11476965;
 RA Lapierre Rodziach K., Rodziach M., Knoll M., Kielar D., Probst M.,
 RA Kiec B., Andrikovics H., Boettcher A., Hubacek J., Aslanidis C.,
 RA Suttorp N., Schmitz G.;
 RA "Homogeneous assay based on 52 primer sets to scan for mutations of
 RT the ABCA1 gene and its application in genetic analysis of a new
 RT patient with familial high-density lipoprotein deficiency syndrome";
 RL Biochim. Biophys. Acta 1537:42-48(2001).
 RN [16]
 RP VARIANTS ID ASN-1289 AND TRP-2981, AND VARIANT IYS-219;
 RX MEDLINE 21464433; PubMed 11476965;
 RA Huang W., Moriyama S., Koda I., Hsu H., Arita M., Kawabata S.,
 RA Kawabata K., Imamura T., Ito T., Kawamura M., Teramoto T., Sasaki J.;
 RA "Novel mutations in ABCA1 gene in Japanese patients with tangier
 RT disease and familial high density lipoprotein deficiency with
 RT coronary heart disease";
 RL Biochim. Biophys. Acta 1537:71-78(2001).
 RN [17]
 RP VARIANTS IYS-219; ALA-399; MET-771; PRO-774; ASN-776; ILE-825;
 RP MET-883; ASP-1172; IYS-1587 AND CYS-1731;
 RX MEDLINE 2146447; PubMed 11476965;
 RA Clee S.M., Zwinderman A.H., Engert J.C., Zwarts K.Y.,
 RA Mulder H.G., Poonp K., Takema J.W., van Willard M., van Dam M.,
 RA Hudson T.J., Brooks Wilson A., Genest J., Jr., Kastelein J.J.P.,
 RA Hayden M.K.;

RT "Common genetic variation in ABCA1 is associated with altered
 RT lipoprotein levels and a modified risk for coronary artery disease";
 RL Circulation 103:1198-1205(2001).
 RN [18]
 RP VARIANT ID THE 256, AND VARIANT ATHEROSCLEROSIS ASP-1611;
 RX MEDLINE-21645894; PubMed-11785958;
 RA Nishida Y., Hirano K., Tsukamoto K., Nakano M., Ikeda C., Poonp K.,
 RA Ishihara M., Sakano N., Zhou Z., Tsujii Y., Matsuyama A., Ohama T.,
 RA Matsuyama F., Ishigami M., Sakai N., Hiraoka H., Hattori H.,
 RA Wellington C., Yoshida Y., Misaki S., Hayden M.K., Egashira T.,
 RA Yamashita S., Matsuzawa Y.;
 RA "Expression and functional analysis of novel mutations of ABCA1-binding
 RT cassette transporter-1 in Japanese patients with high-density
 RT lipoprotein deficiency";
 RL Biochem. Biophys. Res. Commun. 290:713-721(2002).
 CC -1- FUNCTION: CAMP-DEPENDENT AND SULFONYLUREA-SENSITIVE ANION
 CC TRANSPORT. KEY CATEPHERP INFLUENCING INTRACELLULAR CHOLESTEROL
 CC TRANSPORT.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED, BUT MOST ABUNDANT IN
 CC MACROPHAGES.
 CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
 CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
 CC ATP BINDING CASSETTE (ABC) DOMAIN.
 CC -1- DISEASE: DEFECTS IN ABCA1 ARE A CAUSE OF HIGH DENSITY LIPOPROTEIN
 CC DEFICIENCY TYPE I (HDL1), ALSO KNOWN AS TANGIER DISEASE (TD). TD
 CC IS A RECESSIVE DISORDER CHARACTERIZED BY ABSENCE OF HIGH DENSITY
 CC LIPOPROTEIN (HDL) CHOLESTEROL FROM PLASMA, HEPATOSPLENOMEGALY,
 CC PERIPHERAL NEUROPATHY, AND FREQUENTLY PREMATURE CORONARY ARTERY
 CC DISEASE (CAD).
 CC -1- DISEASE: Defects in ABCA1 are a cause of high density lipoprotein
 CC deficiency.
 Query Match 71 19; Score 42; DP 1; Length 2261;
 Best Local Similarity 71.48; Fred. No. 2.1e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NLFETPI 7
 DB 808 NIFESPV 814
 RESULT 15
 ID ABC1_MOUSE STANDARD; PPT: 2261 AA.
 AC P41233;
 DT 01 FEB 1995 (Rel. 31, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DI 01-MAR-2002 (Rel. 41, Last annotation update)
 DE ABC-binding cassette, sub-family A, member 1 (ABC-binding cassette
 DE transporter 1) (ABC-binding cassette 1) (ABC-1).
 GN ABC1 OR ABC1
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurophathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 PP SEQUENCE FROM N.A.
 RX STEIN-DEA/2; Tissue-Macrophage;
 RX MEDLINE-94375008; PubMed-909792;
 RA Luciano M.F., Denicot F., Savary S., Mattei M.-G., Chimini G.;
 RA "Cloning of two novel ABC transporters mapping on human chromosome
 RT 9";
 RL Genomics 21:150-159(1994).
 RN [2]
 PP SEQUENCE FROM N.A.
 RX STRAIN-C57BL/6J;
 RA Qiu Y., Cavellier L., Chiu S., Rubin E., Cheng J.-F.;
 RT "Human and mouse ABCA1 comparative sequencing and transgenesis studies
 RL identify potential regulatory sequences";
 RL Submitted (12-2000) to the EMBL/GenBank/DDBJ databases.
 CC -1- FUNCTION: CAMP-DEPENDENT AND SULFONYLUREA-SENSITIVE ANION
 CC TRANSPORT. KEY CATEPHERP INFLUENCING INTRACELLULAR CHOLESTEROL
 CC TRANSPORT (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN ADULT TISSUES. HIGHEST

CC LEVELS ARE FOUND IN PREGNANT UTERUS AND UTERUS.
 CC -1- DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES.
 CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
 CC ATP BINDING CASSETTE (ABC) DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X75926; CAA53530.1; ALT_INIT.
 CC EMBL: AF287263; AAG39073.1; ALT_INIT.
 CC MGD: MGI:99607; Abcal.
 CC InterPro: IPR003593; AAA.
 CC InterPro: IPR003439; ABC_transportr.
 CC InterPro: IPR001687; ATP_GTP_A.
 CC InterPro: IPR003838; DUF214.
 CC InterPro: IPR000897; SRP54.
 CC Pfam: PF000095; ABC_tran; 2.
 CC Pfam: PF02687; DUF214; 1.
 CC Pfam: PF00448; SRP54; 1.
 CC SMART: SM00382; AAA; 1.
 CC PROSITE: PS00211; ABC_TRANSPORTER; 1.
 CC KW ATP-binding; Glycoprotein; Transmembrane; Transport.
 FT TRANSMEM 26 42 POTENTIAL.
 FT TRANSMEM 640 656 POTENTIAL.
 FT TRANSMEM 690 706 POTENTIAL.
 FT TRANSMEM 717 733 POTENTIAL.
 FT TRANSMEM 749 765 POTENTIAL.
 FT TRANSMEM 771 787 POTENTIAL.
 FT TRANSMEM 1041 1057 POTENTIAL.
 FT TRANSMEM 1351 1367 POTENTIAL.
 FT TRANSMEM 1651 1677 POTENTIAL.
 FT TRANSMEM 1708 1724 POTENTIAL.
 FT TRANSMEM 1737 1753 POTENTIAL.
 FT TRANSMEM 1775 1791 POTENTIAL.
 FT TRANSMEM 1854 1870 POTENTIAL.
 FT NP_BIND 933 940 ATP (POTENTIAL).
 FT NP_BIND 1946 1953 ATP (POTENTIAL).
 FT CARBOHYD 14 14 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 151 151 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 161 161 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 244 244 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 292 292 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 337 337 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 349 349 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 478 478 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 489 489 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 521 521 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 820 820 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1144 1144 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1294 1294 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1453 1453 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1499 1499 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1504 1504 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1637 1637 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2044 2044 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 2238 2238 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 1567 1568 MISSING (IN REF. 2).
 FT CONFLICT 2024 2024 MISSING (IN REF. 2).
 SQ SEQUENCE 2261 AA; 254011 MW; FA62B21F3D509F9 CRG64;

Query Match 71.1%, Score 32, DB 1, Length 2261;
 Best local Similarity 71.4%; Pred. No. 2,1e+02;
 Matches 5; Conservative 2; Mismatches 0, indels 0, Gaps 0;

QY 1 NLEPPI 7
 Db 808 NLESPV 814

Search completed: September 5, 2002, 15:31:41
 Job time: 476 sec

GenCore version 4.5
Copyright (c) 1993 2000 Compugen Ltd

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:31:01, Search time 122.86 seconds
(without alignments)
12.673 Million cell updates/sec

Title: US-09-744-804-37
Perfect score: 45
Sequence: 1 NLEFPIA 9

Scoring table: BLOSUM62
Gapop 10 0, Gapext 0 5

Searched: 56222 seqs, 17294939 residues

Total number of hits satisfying chosen parameters: 62222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post processing: Minimum Match 84
Maximum Match 100%
Listing first 45 summaries

Database: SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_muc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_prodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	100.0	335	4 Q9RTL9	Q9RTL9 homo sapien
2	39	84.7	363	6 Q97718	Q97718 chlamydia
3	36	80.0	541	17 Q97282	Q97282 schistosom
4	36	80.0	669	10 Q9L227	Q9L227 arabidopsis
5	34	80.0	700	13 Q9L227	Q9L227 arabidopsis
6	34	75.6	250	16 Q96544	Q96544 mycobacteri
7	34	75.6	1244	2 Q9PAH3	Q9PAH3 noster sp.
8	34	73.3	245	15 Q95814	Q95814 human immun
9	33	73.3	304	16 Q92AH5	Q92AH5 listeria in
10	33	73.3	572	5 Q44568	Q44568 caenorhabdi
11	33	73.3	658	3 Q74798	Q74798 schistosom
12	33	73.3	654	5 Q21235	Q21235 caenorhabdi
13	33	73.3	1162	4 Q9RV54	Q9RV54 homo sapien
14	33	73.3	1241	4 Q9Y4B4	Q9Y4B4 homo sapien
15	33	73.3	1466	11 Q9QNG0	Q9QNG0 mus muscul
16	33	73.3	2205	5 Q9NG02	Q9NG02 dictyostell

17	32	71.1	78	4 P78438	P78438 homo sapien
18	32	71.1	121	17 Q58495	Q58495 pyrococcus
19	32	71.1	173	16 Q9P724	Q9P724 vibrio chol
20	32	71.1	208	17 Q59231	Q59231 pyrococcus
21	32	71.1	245	16 Q9A1C1	Q9A1C1 streptococ
22	32	71.1	255	2 Q9AHV1	Q9AHV1 pseudomonas
23	32	71.1	255	16 Q91010	Q91010 pseudomonas
24	32	71.1	268	16 Q92712	Q92712 rhizobium m
25	32	71.1	278	16 Q9ABA4	Q9ABA4 caulobacter
26	32	71.1	311	16 Q9BHW3	Q9BHW3 rhizobium l
27	32	71.1	328	15 Q9EG03	Q9EG03 human immun
28	32	71.1	328	15 Q9EG03	Q9EG03 human immun
29	32	71.1	333	5 Q9VNR0	Q9VNR0 drosophila
30	32	71.1	394	16 Q9PED1	Q9PED1 xylella fas
31	32	71.1	398	16 Q9HW26	Q9HW26 pseudomonas
32	32	71.1	399	10 Q92554	Q92554 arabidopsis
33	32	71.1	399	16 Q9H0P4	Q9H0P4 pseudomonas
34	32	71.1	417	3 Q9H1D0	Q9H1D0 schistosom
35	32	71.1	419	12 P87652	P87652 equine herp
36	32	71.1	425	16 Q9PW06	Q9PW06 chlamydia m
37	32	71.1	435	16 Q93357	Q93357 bacillus ha
38	32	71.1	445	2 Q52293	Q52293 agrobacteri
39	32	71.1	456	10 Q9FE23	Q9FE23 arabidopsis
40	32	71.1	530	16 Q92559	Q92559 mycobacteri
41	32	71.1	586	1 Q9C4P0	Q9C4P0 halobacteri
42	32	71.1	590	2 Q57154	Q57154 acinetobact
43	32	71.1	670	10 Q9Z0M7	Q9Z0M7 arabidopsis
44	32	71.1	779	3 Q12168	Q12168 saccharomye
45	32	71.1	879	16 P71076	P71076 bacillus su

ALIGNMENTS

RESULT 1

ID Q9RTL9 PRELIMINARY: PRI: 335 AA.
AC Q9RTL9
DT 01-JUN-2001 (TRMBLrel, 17, Created)
DT 01-JUN-2001 (TRMBLrel, 17, Last sequence update)
DT 01-DEC-2001 (TRMBLrel, 19, Last annotation update)
DE SIMILAR TO MILK FAT GLOBULE-EGF FACTOR 8 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota, Metazoa, Chordata, Eumetazoa, Vertebrata, Euteleostomi;
OC Mammalia, Eutheria, Primates, Catarrhini, Hominoidea, Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
EC TISSUE: MELANOMA;
RA Strausberg R.;
FL Submitted (11/1/2001) to the EMBL/GenBank/DDBJ databases.
CC -1- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.
DR EMBL; BC003610; AA003610.1; -;
DR HSSP; P08709; 1BF9.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001438; EGF_H1.
DR InterPro; IPR000421; FA58_C.
DR PRINTS; PR00010; EGFRI000.
DR SMART; SM00191; EGF_1.
DR SMART; SM00231; EGF-like; 1.
DR SMART; SM00231; FA58C_2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01186; FA58C_1; 1.
DR PROSITE; PS01285; FA58C_1; 1.
DR PROSITE; PS01286; FA58C_2; 2.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 335 AA, 37523 MW, 2068451878025 CRC64;

Query Match 100.0% Score 45; DB 4; Length 335;
Best Local Similarity 100.0%; Pred No. 0.19;
Matches 9, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

```

QY 1 NLFETPILA 9
DB 304 NLFETPILA 312

RESULT 2
QY7718 PRELIMINARY: PRT: 363 AA.
AC 07718;
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DI 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DE SPERM-MEMBRANE ASSOCIATED PROTEIN P47 (FRAGMENT).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi.
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID: 9706;
RN 111
RP SEQUENCE FROM N.A.
RC TISSUE: TESTIS;
RA Gestel M., Toepfer-Petersen E.;
RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases
DR EMBL: AJ010121; CAA09010.1; -
DR HSSP: P00740; IEDM.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF_1.
DR Pfam: PF00754; P5_P8_type_C; 2.
DR SMART: SM00181; EGF_1.
DR SMART: SM00241; FAS9C; 2.
DR PROSITE: PS00022; EGF_1; UNKNOWN_2.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS01285; FAS9C_1; 1.
KW EGF-like domain; Glycoprotein.
FT NON_TER 1
FT NON_TER 363
SQ SEQUENCE 463 AA: 40744 MW: 1F8B6395AF42338D CRC64;

Query Match: 86.7%; Score 36; DB 6; Length 363.
Best Local Similarity 77.8%; Pred. No. 3.9;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLFETPILA 9
DB 339 NMFETPILA 347

RESULT 3
QY72X2 PRELIMINARY: PRT: 531 AA.
AC 072X2;
DI 01-DEC-2001 (Tremblrel. 19, Created)
DE 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN ST1017.
GN ST1017.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID: 111955;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN JCM 10545 / 7;
RX PubMed: 11572479;
RA Kawababayasi Y., Hino Y., Morikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Aoki A., Kusagi H., Hosoyama A., Fukui S.,
RA Nakai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshikawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Ouchi A.,
RA Aoki K.-I., Masuda S., Yamanli M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).

DR EMBL: AP000984; BAB66041.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 541 AA: 58083 MW: 75R3PQ3B0H4165R CRC64;

Query Match: 80.0%; Score 36; DB 17; Length 541.
Best Local Similarity 66.7%; Pred. No. 26;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLFETPILA 9
DB 102 NLFETPILA 110

RESULT 4
QY1227 PRELIMINARY: PRT: 669 AA.
AC 091227;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DI 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PYP-TIPTICAI 75.3 KDA PROTEIN (GENOMIC DNA, CHROMOSOME 5, P1
DE CLONE:MOK11).
GN TIE3_60.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids 11; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID: 3702;
RN 111
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardiles W., Bayssacert G., Dasseville R.,
RA De Clerck K., De Keyser A., Noyt P., Rouze P., Van Den Daele H.,
RA Villarroel R., Gieles J., Van Montagu M., Bancroft I., Meuwis H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN 121
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RA Schmitz (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN 131
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE-98162728; PubMed 9501997;
PA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
PA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5, III.
RT Sequence features of the regions of 1,191,918 bp covered by seventeen
RT physically assigned P1 clones.";
RL DNA Res. 4:401-414(1997).
DR EMBL: AL162972; CAB86012.1; -
DR EMBL: AB029271; BAB08974.1; -
DR InterPro: IPR002110; ANK.
DR PROSITE: PS00297; ANK_REPEAT_REGION; 1.
KW ANK repeat; Hypothetical protein; Repeat.
SQ SEQUENCE 669 AA: 75339 MW: 676AC37958851B4 CRC64;

Query Match: 80.0%; Score 36; DB 10; Length 669;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPILA 8
DB 112 NLFETPILA 119

RESULT 5
QY1229 PRELIMINARY: PRT: 705 AA.
AC 091229;
DI 01-OCT-2000 (Tremblrel. 15, Created)
DE 01-OCT-2000 (Tremblrel. 15, Last sequence update)

```

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYDROTHERMAL 79.5 KDA PROTEIN.
 GN TLE3_40.
 OS Arabidopsis thaliana (Mouse ear cress).
 OC Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,
 OC Spermatophyta, Magnoliophyta, Eudicotyledons, Core eudicots, Rosidae,
 OC eustosids II, Brassicales, Brassicaceae, Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 FA Bevan M., Terry N., Ardiles W., Boysschaert C., Dassocheville R.,
 FA de Clerck P., De Knyser A., Nuyt P., Ponce F., Van den Daele H.,
 FA Villarejo P., Gielens J., Van Montagu M., Bancroft I., Mowes H.W.,
 RA Rudd S., Lemcke K., Meyer K.F.X.,
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project.
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL162972; CAB6010.1;
 DR InterPro: IPR002110; ANK.
 DR PROSITE: PS0297; ANK_PEP_REGION; 1.
 KW ANK repeat; Hypothetical protein; Repeat.
 SQ SEQUENCE 705 AA; 79491 MW; 91c44e144f012dc7 CRC64;

Query Match 80.0%; Score 66; PB 10; Length 705;
 Best Local Similarity 75.0%; Pred No 34;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLETPIL 8
 |||:|:|
 DB 99 NLEYTPLL 106

RESULT 6
 ID Q06544 PRELIMINARY; PRT; 250 AA.
 AC Q06544;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE FF-FEABLE 3 HYDROXYACYL-CoA DEHYDROGENASE 25.8 KDA TYPE II
 DE (EC 1.1.1.35) (TYPE II HADH)
 GN RV1144 OR MTC165.11
 OS Mycobacterium tuberculosis.
 CC Pathogen; Filicidal; Actinobacteria; Actinobacteriales;
 CC Actinobacteriales; Mycobacteriaceae; Mycobacterium; Mycobacterium
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 FA Cole S.T., Brosch P., Parkhill J., Garnier T., Churcher C., Harris D.,
 FA Cado-Ciro S.V., Eiglacik K., Ras S., Berry S.E., Hill J., Leclaire F.,
 RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 FA Harty T., Jangals K., Krash A., Mclennan T., Muthy S., Murphy L.,
 RA Oliver S., Ostroff S., Quail M.A., Rajagubalan M.A., Rupp S.T.,
 RA Putter S., Seeger K., Skellern S., Squares S., Squares P.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence";
 PL Nature 393:537-544(1998).
 CC -1- CATALYTIC ACTIVITY: L-3-HYDROXYACYL-CoA + NAD(+) = L-3-OXOACYL-CoA +
 NADH.
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/DEHYDROGENASES
 CC FAMILY (SDP).
 DR FMRU: 295584; CAB09032.1;
 DR HSP: 070351; 183S
 DR Tuberculid; RV1144;
 DR InterPro: IPR02198; ANK_Short;

DR InterPro: IPR002055; NAD_binding.
 DR Pfam: PF00106; adh_short; 1.
 DR PROSITE: PS00061; ADH_Short; 1.
 KW Hypothetical protein, Oxidoreductase, NAD; Mitochondrion;
 KW Eukaryotic Proteome.
 FT DOMAIN 32 35 POLY-VAL.
 FT NGBIND 7 32 NAD (BY SIMILARITY).
 FT ACT_SITE 156 156 BY SIMILARITY.
 SQ SEQUENCE 250 AA; 25787 MW; 12418b5a19c4e4 CRC64;

Query Match 75.0%; Score 34; PB 10; Length 250;
 Best Local Similarity 75.0%; Pred No 32;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LNETPILA 9
 |||:|:|
 DB 188 LNETPILA 195

RESULT 7
 ID Q9RAH3 PRELIMINARY; PRT; 1244 AA.
 AC Q9RAH3;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE NCSR.
 GN NOSTOC.
 OS Nostoc sp. GSV224.
 CC Bacterial; Cyanobacteria; N-Stinales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=76334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GSV224;
 RA Hoffmann D., Hevel J.M., Moore K.E.;
 PT "Characterization of the nostocopeptide biosynthetic gene cluster of
 Nostoc sp. GSV224";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF204805; AAF1592.2;
 DR HSP: P25715; NMLA.
 DR InterPro: IPR01227; Acyltransferase;
 DR InterPro: IPR00754; Poly-yl-ylase;
 DR InterPro: IPR01993; Mitochondrion;
 DR InterPro: IPR03880; Phosphatase;
 DR Pfam: PF00698; Acyltransferase;
 DR Pfam: PF0107; ketocyclin; 1.
 DR Pfam: PF02881; ketocyclin; 1.
 DR Pfam: PF00550; pp-binding; 1.
 DR PROSITE: PS00075; ACP_DOMAIN; 1.
 DR PROSITE: PS00095; K-KEP-ACYL SYNTHASE; UNFOWN_1.
 DR PROSITE: PS00315; MITOCHONDRIAL CARRIER; UNFOWN_1.
 DE FE001E; PS00315; PHOSPHATETHEINE; UNFOWN_1.
 KW Phosphatetheine.
 SQ SEQUENCE 1244 AA; 127519 MW; A9CE742E6A8A81 CRC64;

Query Match 75.0%; Score 34; PB 2; Length 1244;
 Best Local Similarity 66.74%; Pred No 17502;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLETPILA 9
 |||:|:|
 DB 1210 NLETPILA 1218

RESULT 8
 ID Q75814 PRELIMINARY; PRT; 245 AA.
 AC Q75814;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

```

DE REVERSE TRANSCRIPTASE (FRAGMENT).
GN Pol.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviridae; Lentivirus.
CX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN TP324 FROM AUSTRALIA;
RA Clad N.R., Hallett L., Neilan B.A., Cooper D.A., Delany S.F.,
RA McQueen P.W.;
RA "Sequence analyses of the reverse transcriptase region of HIV-1
RA isolates from Sydney, Australia.";
RA Submitted (JUL 1996) to the EMBL/GenBank/DBJ databases.
RL EMBL: U64171; AAB05314.1; -.
DR InterPro: IPR000477; RVTSE.
DR Pfam: PF00078; JVT; 1.
KW RNA-directed DNA polymerase.
FT NON_TER 1
FT NON_TER 245
SQ SEQUENCE 245 AA; 28756 MW; 6D51EAA38F6480CB664;
Query Match 73.3%; Score 33; DB 15; Length 245;
Best Local Similarity 66.7%; Prod No 51;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LFETPILA 9
DB 44 MYETPILA 50
|||||

RESULT 9
Q92AH5 Q92AH5 PRELIMINARY; PRI: 304 AA.
AC Q92AH5;
DT 01-DEC-2001 (TEMBLrel 19, Created)
DT 01-DEC-2001 (TEMBLrel 19, Last sequence update)
DT 01-DEC-2001 (TEMBLrel 19, Last annotation update)
DE PYRD PROTEIN.
DR InterPro: IPR000197.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacilli; Clostridium group.
CX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-CLIP 11562 / SHC-VAP 5A;
RA PubMed:11679669;
RA Glaser P., Frangeul L., Buchrieset C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Couye E., de Barrovar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Esibi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapov G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjati H.,
RA Nordstedt G., Novilla S., de Pablos R., Perez Diaz J.-C., Putcell R.,
RA Remmel B., Rose M., Schlueder T., Simoes N., Tierney A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Weissart P.;
RA "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001)
DR EMBL: AL596170; CAC91777.1; -.
DR DDBJ/EMBL/GenBank: L1N01947; -.
KW Complete proteome
SQ SEQUENCE 304 AA; 32225 MW; 6736141CA4B8291CB664;
Query Match 73.3%; Score 33; DB 16; Length 304;
Best Local Similarity 85.7%; Prod No 64;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LFETPILA 9
DB 44 MYETPILA 50
|||||

```

```

DB 92 LFETPILA 98
RESULT 10
O44508 O44508 PRELIMINARY; PRI: 572 AA.
AC O44508;
DT 01-JUN-1998 (TEMBLrel 06, Created)
DT 01-JUN-1998 (TEMBLrel 06, Last sequence update)
DT 01-JUN-2001 (TEMBLrel 19, Last annotation update)
DE HYPOPHYSICAL 66.6 KDA PROTEIN.
GN F42G8.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Phasmatodea; Peleodermidae; Caenorhabditis.
CX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RA MEDLINE=9409415; PubMed=9451916;
RA None;
RA "Genome sequence of the nematode C. elegans, a platform for
RA investigating biology, the C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RA Gattung S., Holmes A.;
RA "The sequence of C. elegans cosmid F42G8.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN-BRISTOL N2;
RA Waterston R.;
RA "Direct Submission.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DE EMBL: AF039618; AAB92073.1; -.
DR InterPro: IPR001763; Fibronase-domain.
DR InterPro: IPR000063; Thiores.
DR SMART: SM00450; RHOD; 1.
KW Hypothetical protein.
SQ SEQUENCE 572 AA; 66648 MW; 5B1081F2343874C0CB64;
Query Match 73.3%; Score 33; DB 5; Length 572;
Best Local Similarity 75.0%; Prod No 120-02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LFETPILA 9
DB 455 VFETPILA 462
|||||

RESULT 11
Q74798 Q74798 PRELIMINARY; PRI: 658 AA.
AC Q74798;
DT 01-NOV-1998 (TEMBLrel 08, Created)
DT 01-NOV-1998 (TEMBLrel 08, Last sequence update)
DT 01-001-2000 (TEMBLrel 15, Last annotation update)
DE HYPOPHYSICAL 57.9 KDA PROTEIN CYH10.04 IN CHROMOSOME II.
GN SPBC2D10.04.
OS Schizosaccharomyces pombe (fission yeast).
CX Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
CX Schizosaccharomycetes;
CX Schizosaccharomycetes;
CX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=372;
RA Wood V., Rajandream M.A., Barrell R.G., Taylor K., Harris D.;
RA Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO YEAST YJL084L.

```

DR EMBL: AL031788; CAA21162.1; 49PR0658P970B4F7 CXC64;
SQ SEQUENCE 658 AA; 72741 MW; 49PR0658P970B4F7 CXC64;

Query Match 73.3%; Score 33; DB 3; Length 618;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPIL 6
||| |||
DB 423 NLFETPIL 428

RESULT 12
Q21225
ID C21225 PRELIMINARY; PPT; 694 AA.

AC Q21225
DT 01-NOV-1996 (TREMURel. 01, Created)
DT 01-NOV-1996 (TREMURel. 01, Last sequence update)
DT 01-DEC-2001 (TREMURel. 19, Last annotation update)
DE K0462.6 PROTEIN.

GN K0462.6.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea, Phallozoa, Phallozoa, Phallozoa.

OC Eubacteriidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]_TaxID=6239;

RP SEQUENCE FROM N.A.

RA Gardner A.E.;

RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA none;

RT "Genome sequence of the nematode C. elegans: A platform for

RT investigating biology."

RL Science 282:2012-2018(1998).

DP EMBL: 775712; CAB00043.1; -

SQ SEQUENCE 694 AA; 78366 MW; 9F465P087A3D49P CXC64;

Query Match 73.3%; Score 33; DB 5; Length 694;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLFETPIL 9
||| |||
DB 584 NLVESPIA 592

RESULT 13
Q9BV54
ID Q9BV54 PRELIMINARY; PPT; 1162 AA.

AC Q9BV54
DT 01-JUN-2001 (TREMURel. 17, Created)
DT 01-JUN-2001 (TREMURel. 17, Last sequence update)
DT 01-DEC-2001 (TREMURel. 19, Last annotation update)
DE HYPOTHETICAL 128.7 KDA PROTEIN (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=PLACENTA, CHORIOCARCINOMA;

RA Strausberg R.;

RL Submitted (DEC 2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: RC001474; AA001474.1; -

DR InterPro: IPR001410; Pfam

DR InterPro: IPR001410; Helicase_C.

DR Pfam: PF00271; Helicase_C.1.

DR Pfam: PF00176; SNF2_N; 1

DR Pfam: PF00176; SNF2_N; 1

DR SMART: SM00487; DEXDC; 1;
DR SMART: SM00490; HELICE; 1;
KW ATP-binding; Helicase; Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 1162 AA; 128676 MW; 42FE1C6114B7N10 CXC64;

Query Match 73.3%; Score 33; DB 4; Length 1162;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLFETPIL 8
||| |||
DB 217 NMFERPIL 224

RESULT 14

Q9Y4B4

ID Q9Y4B4 PRELIMINARY; PPT; 1241 AA.

AC Q9Y4B4

DT 01-NOV-1999 (TREMURel. 12, Created)

DT 01-NOV-1999 (TREMURel. 12, Last sequence update)

DT 01-DEC-2001 (TREMURel. 19, Last annotation update)

DE KIAA0809.

GN KIAA0809.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=BRAIN;

EX MEDLINE=99087487; PubMed=9872452;

RA Nakase T., Ishikawa K., Suyama M., Kikano R., Miyajima N., Tanaka A.,

RA Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XI.

RT The complete sequences of 100 new cDNA clones from brain which code

RT for large proteins in vitro."

RL DNA Res. 5:277-286(1998).

DR EMBL: AB018352; BAA34529.1; -

DR InterPro: IPR001410; DEAD.

DR InterPro: IPR001650; Helicase_C.

DR Pfam: PF00271; Helicase_C.1.

DR Pfam: PF00176; SNF2_N; 1.

DR SMART: SM00487; DEXDC; 1.

DR SMART: SM00490; HELICE; 1.

KW ATP-binding; Helicase.

FT NON_TER 1

SQ SEQUENCE 1241 AA; 137451 MW; E7H1IA45P8769C40 CXC64;

Query Match 73.3%; Score 33; DB 4; Length 1241;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLFETPIL 8
||| |||
DB 296 NMFERPIL 303

RESULT 15

Q99NGO

ID Q99NGO PRELIMINARY; PPT; 1466 AA.

AC Q99NGO

DT 01-JUN-2001 (TREMURel. 17, Created)

DT 01-JUN-2001 (TREMURel. 17, Last sequence update)

DT 01-DEC-2001 (TREMURel. 19, Last annotation update)

DE SHERIDAN RECEPTOR-INTERACTING SNF2 DOMAIN PROTEIN.

GN SRSN2L OR SRSN2L.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

```

>X NCBI_TaxID=10090;
>N [1]
>P SEQUENCE FROM N.A.
>C STRAIN SWISS WEBSTER /NIH; TISSUE=11-DAY EMBRYO;
>R Rouleau N., Reeben M., Mollanen A.M., Palvimo J.T., Janne O.A.;
>A "Identification of a novel SNF2 like family member protein as a
>T coregulator in steroid receptor-mediated gene transcription.";
>L Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
>D EMBL; AJ132389; CAC24703.1; -.
>D MGD; MG11933196; Srisu121.
>D InterPro: IPR001410; DEAD.
>D InterPro: IPR001650; Helicase_C.
>D InterPro: IPR000430; SNF2_N.
>D Pfam: PF00271; helicase_C; 1.
>D Pfam: PF00176; SNF2_N; 1.
>D SMART: SM00487; DEXDC; 1.
>D SMART: SM00490; HELICE; 1.
>K ATP-binding; Helicase; Receptor.
>S SEQUENCE 1466 AA; 162539 MW; D5945AL802B03012 CERC4;

```

```

Query Match          74.8%; Score 33; DB 11; Length 1466;
Best Local Similarity 75.0%; Prod. No. 3.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 NIEFPHL 8
Db 521 NIEFPHL 528

```

```

Search completed: September 5, 2002, 15:31:03
Job time: 468 sec

```


GenGene version 4.5
Copyright (c) 1993 - 2000 Compugen Inc

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:27:34 : Search time 58.96 Seconds
(without alignments)
3.735 Million cell updates/sec

Title: us-09-744-804-38
Perfect score: 46
Sequence: 1 NLETPVEA 9

Scoring table: BLOSUM62
Gapop 10 0, Gapext 0 5

Searched: 23158 seqs, 2445594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA**
1: /cgn2-6/ptdata/2/1aa/5A.COMR.pcp.*
2: /cgn2-6/ptdata/2/1aa/5B.COMR.pcp.*
3: /cgn2-6/ptdata/2/1aa/6A.COMR.pcp.*
4: /cgn2-6/ptdata/2/1aa/6B.COMR.pcp.*
5: /cgn2-6/ptdata/2/1aa/6C.COMR.pcp.*
6: /cgn2-6/ptdata/2/1aa/6D.COMR.pcp.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	46	100.0	160	2	US-08-162-402B-10
2	46	100.0	217	1	US-07-637-538-3
3	46	100.0	217	2	US-08-162-402B-3
4	46	100.0	218	1	US-07-607-638C-2
5	46	100.0	218	2	US-08-162-402B-2
6	46	100.0	487	2	US-08-162-402B-6
7	46	100.0	465	2	US-08-162-402B-9
8	37	80.4	170	2	US-08-162-402B-12
9	36	78.3	14	2	US-08-162-402B-26
10	32	69.6	406	4	US-08-584-000A-8
11	31	67.4	108	3	US-04-187-859-40
12	31	67.4	152	3	US-08-936-165A-344
13	31	67.4	217	3	US-08-971-178-1
14	31	67.4	217	3	US-08-971-158-3
15	30	65.2	127	4	US-08-437-054A-12
16	30	65.2	491	4	US-08-984-618-11
17	30	65.2	739	4	US-04-136-320-6
18	30	65.2	867	4	US-08-659-261-2
19	30	65.2	867	4	US-09-296-490-2
20	30	65.2	857	5	PCT-US96-11445-2
21	30	65.2	1678	4	US-08-932-135-22
22	29	63.0	68	3	US-08-700-152A-7
23	29	63.0	68	3	US-08-912-272-5
24	29	63.0	68	4	US-09-026-039-5
25	29	63.0	79	4	US-08-904-263A-6
26	29	63.0	181	4	US-08-858-207A-41
27	29	63.0	188	2	US-08-885-500A-6

Sequence 28, Appl
Sequence 6, Appl
Sequence 28, Appl
Sequence 10, Appl
Sequence 4, Appl
Sequence 4, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 44, Appl
Sequence 2, Appl
Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-162-402B-10
: Sequence 10, Application US/0816-402B
: Patent No. 5972337

: GENERAL INFORMATION:
: APPLICANT: GERTANI, ROBERTO L.
: APPLICANT: PETERSON, JERRY A.
: APPLICANT: LAPOCCA, DAVID J.
: TITLE OF INVENTION: 45 KILTON HUMAN MILK FAT
: TITLE OF INVENTION: CLOSTRIDIAL GLOBULE (HMG) ANTIGEN, FRAGMENTS & PEPTIDE PROTEIN
: NUMBER OF EXPERIMENTS: 29
: CORRESPONDENT ADDRESS:
: ADDRESSEE: Fritty, Schaefer & Popiawski
: STREET: 444 South Flower St., 19th Floor
: CITY: Los Angeles
: STATE: CA
: COUNTRY: USA
: ZIP: 90071

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPL. FILE NAME: US-08-162-402B
: FILING DATE: 03-DEC-1993
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Azzel, Viviana
: REGISTRATION NUMBER: 30,930
: TELEPHONE: 213-622-7700
: TELEFAX: 213-489-4210
: TELEX:
: INFORMATION FOR SEQ ID NO: 10:

: SEQUENCE CHARACTERISTICS:
: LENGTH: 160 amino acids
: TYPE: amino acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: peptide
: US 08 162 402B 10

Query Match 100.0% Score 46. DB 2 Length 160;
Best Local Similarity 100.0% Pred. No. 0.038;

```

Matches 9, Conservative 0, Mismatches 0, Indels 0, Gaps 0
QY 1 NLPETPVEA 9
    | | | | | | | |
DB 126 NLPETPVEA 134

RESULT 2
US-07-607-538C-3
: Sequence 3, Application US/07607538C
: Patent No. 5455031
: GENERAL INFORMATION:
: APPLICANT: Ceriani Dr., Roberto L.
: APPLICANT: Peterson Dr., Jerry A.
: APPLICANT: Larocca, David J.
: TITLE OF INVENTION: POLYPEPTIDE WITH 46
: TITLE OF INVENTION: DIFFERENTIATION ANTIGEN BINDING SPECIFICITY AND CLOTTING
: TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGUES,
: TITLE OF INVENTION: FUSION PROTEIN, POLYNUCLEOTIDE AND POLYRIBO-
: TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-
: TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF
: TITLE OF INVENTION: USE THEREOF
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: V. Anzel & Assoc.
: STREET: 2055 No. 5455031th Broadway
: CITY: Walnut Creek
: STATE: California
: COUNTRY: USA
: ZIP: 94596
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS 5.0
: SOFTWARE: Patent to P-1-45-#1 0, Version: #1 25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/607,538C
: FILING DATE: 01-NOV-1990
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Viviana Anzel
: REGISTRATION NUMBER: 30,930
: REFERENCE/DOCKET NUMBER: CPEL-004
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (510) 943-1931
: TELEFAX: (510) 943-1189
: TELEX: N.A.
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 217 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FRAGMENT TYPE:
US-07-607-538C-3

Query Match 100.0%; Score 46, DB 1; Length 217;
Best Local Similarity 100.0%; Pred No. 0.054;
Matches 9, Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 1 NLPETPVEA 9
    | | | | | | | |
DB 24 NLPETPVEA 32

RESULT 3
US-08-162-402B-3
: Sequence 3, Application US/08162402B
: Patent No. 5972337
: GENERAL INFORMATION:
: APPLICANT: CERIANI, ROBERTO L.

```

```

: APPLICANT: PETERSON, JERRY A.
: APPLICANT: LARocca, DAVID J.
: TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
: TITLE OF INVENTION: GLIADIN (HMG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: pretty, Schroeder & Poplawski
: STREET: 444 South Flower St., 19th Floor
: CITY: Los Angeles
: STATE: CA
: COUNTRY: USA
: ZIP: 90071
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/162,402B
: FILING DATE: 03-DEC-1993
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Amzel, Viviana
: REGISTRATION NUMBER: 30,930
: REFERENCE/DOCKET NUMBER: P66 38215
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 213-622-7700
: TELEFAX: 213-489-4210
: TELEX:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 217 amino acids
: TYPE: amino acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: peptide
US-08-162-402B-3

Query Match 100.0%; Score 46; DB 2; Length 217;
Best Local Similarity 100.0%; Pred. No. 0.053;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLPETPVEA 9
    | | | | | | | |
DB 24 NLPETPVEA 32

RESULT 4
US-07-607-538C-2
: Sequence 2, Application US/07607538C
: Patent No. 5455031
: GENERAL INFORMATION:
: APPLICANT: Ceriani Dr., Roberto L.
: APPLICANT: Peterson Dr., Jerry A.
: APPLICANT: Larocca, David J.
: TITLE OF INVENTION: POLYPEPTIDE WITH 46
: TITLE OF INVENTION: DIFFERENTIATION ANTIGEN BINDING SPECIFICITY AND CLOTTING
: TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGUES,
: TITLE OF INVENTION: FUSION PROTEIN, POLYNUCLEOTIDE AND POLYRIBO-
: TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-
: TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF
: TITLE OF INVENTION: USE THEREOF
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: V. Anzel & Assoc.
: STREET: 2055 No. 5455031th Broadway
: CITY: Walnut Creek
: STATE: California
: COUNTRY: USA

```

ZIP: 94596
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 5.0
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/607 538C
FILING DATE: 01-NOV-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Viviana Amzel
REGISTRATION NUMBER: 30,930
REFERENCE/AVCKET NUMBER: P56 38215
TELEPHONE: (510) 943-1931
TELEFAX: (510) 943-1189
TELEX: N.A.
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE:
US-07-607-538C-2

Query Match 100.0% Score 46; DB 1; Length 218;
Best Local Similarity 100.0% Pred. No. 0.053;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 NLFETIVEA 9
DB 25 NLFETIVEA 33

RESULT 5
US-08-162-402B-2
Sequence 2, Application: US/08/162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERIAI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LAROCCA, DAVID J.
TITLE OF INVENTION: 46 KILATON HUMAN MILK FAT
TITLE OF INVENTION: GLOBULE (HMRG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplawski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162.402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/AVCKET NUMBER: P56 38215
TELECOMMUNICATION INFORMATION:

TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-2

Query Match 100.0% Score 46; DB 2; Length 218;
Best Local Similarity 100.0% Pred. No. 0.053;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 NLFETIVEA 9
DB 25 NLFETIVEA 33

RESULT 6
US-08-162-402B-6
Sequence 5, Application: US/08/162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERIAI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LAROCCA, DAVID J.
TITLE OF INVENTION: 46 KILATON HUMAN MILK FAT
TITLE OF INVENTION: GLOBULE (HMRG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplawski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/162.402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/AVCKET NUMBER: P56 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 387 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-6

Query Match 100.0% Score 46; DB 2; Length 387;
Best Local Similarity 100.0% Pred. No. 0.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Caps 0;

```

QY      1 NLFETPVEA 9
      |||||
DB      194 NLFETPVEA 202

RESULT 7
US-08-162-402B-8
; Sequence 8, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CRIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
; TITLE OF INVENTION: GLOBULE (HMEG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 465 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-162-402B 8

Query Match      100.0%, Score 46, DB 2, Length 465,
Best Local Similarity 100.0%, Pred. No. 0.13;
Matches 9; Conservative 6, Mismatches 6, Indels 0, Gaps 0;

QY      1 NLFETPVEA 9
      |||||
DB      272 NLFETPVEA 280

RESULT 8
US-08-162-402B-12
; Sequence 12, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CRIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
; TITLE OF INVENTION: GLOBULE (HMEG) ANTIGEN, FRAGMENTS & FUSION PROTEIN

```

```

; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 159 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-162-402B-12

Query Match      80.4%, Score 37, DB 2, Length 159,
Best Local Similarity 77.8%, Pred. No. 2.4;
Matches 7; Conservative 1, Mismatches 1, Indels 0, Gaps 0;

QY      1 NLFETPVEA 9
      |||||
DB      128 NLFETPVEA 136

RESULT 9
US-08-162-402B-26
; Sequence 26, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CRIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LAROCCA, DAVID J.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
; TITLE OF INVENTION: GLOBULE (HMEG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993

```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: P66 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 14 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B 26

Query Match 78.3%; Score 35; DR 2; Length 14;
Best Local Similarity 75.0%; Pred. No. 0.24;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPVE 8
DB 2 NLFETPVE 9
|||||:|

RESULT 10
US-08-584-008A-8
Sequence 8, Application US/08584008A
Patent No. 6277615
GENERAL INFORMATION:
APPLICANT: Varghese, Joseph N.
APPLICANT: Garrett, Thomas P.J.
APPLICANT: Fincher, Geoffrey B.
APPLICANT: Hoj, Peter B.
APPLICANT: Chen, Lin
TITLE OF INVENTION: BETA-GLUCANASE OF ENHANCED STABILITY
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN
STREET: The J-unit Building, 400 Seventh St N W
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/584.008A
FILING DATE: 11-JAN-1996
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PL9821
FILING DATE: 07-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU94/00377
FILING DATE: 06-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Holman, J.C.
REGISTRATION NUMBER: 22,769
REFERENCE/DOCKET NUMBER: 4444/P660173
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-538-6666
TELEFAX: 202-393-5350
INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 306 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-584-008A-8

Query Match 59.5%; Score 32; DR 4; Length 306;
Best Local Similarity 66.7%; Pred. No. 51;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLFETPVEA 9
DB 205 NLFETPVEA 213
|||||:|

RESULT 11
US 09 187 859-40
Sequence 40, Application US/09187859A
Patent No. 6358920
GENERAL INFORMATION:
APPLICANT: Blaschuk, Orest W.
APPLICANT: Gour, Barbara J.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR MODULATING NONCLASSICAL
TITLE OF INVENTION: CADHERIN-MEDIATED FUNCTIONS
FILE REFERENCE: 100086.407C1
CURRENT APPLICATION NUMBER: US/09/187.859A
CURRENT FILING DATE: 1998-11-06
NUMBER OF SEQ ID NOS: 4052
SOFTWARE: Patent In Ver 2 0
SEQ ID NO 40
LENGTH: 108
TYPE: PRT
ORGANISM: Mus musculus
US-09-187-859-40

Query Match 67.4%; Score 31; DR 4; Length 108;
Best Local Similarity 75.0%; Pred. No. 25;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFEPTVEA 9
DB 73 LYEIPVEA 80
|-|:|:|

RESULT 12
US-08-936-165A-344
Sequence 344, Application US/08936165A
Patent No. 6348582
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Burnham, Martin
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Lovett, Michael
APPLICANT: Nicholas, Richard
APPLICANT: Pratt, Julie
APPLICANT: Reichard, Richard
APPLICANT: Rosenberq, Martin
APPLICANT: Ward, Judith
TITLE OF INVENTION: No. 6348582-1 Polypeptide in Polypeptide Libraries,
TITLE OF INVENTION: Polypeptides and Their Uses
NUMBER OF SEQUENCES: 534
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/44,875A
FILING DATE: 24-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/827,032
FILING DATE: 24-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmli, Edward R.
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: PF-0245-1
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 344:
SEQUENCE CHARACTERISTICS:
LENGTH: 162 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-936-165A-344

Query Match 67.4%, Score 31, DB 3, Length 162;
Best Local Similarity 62.5%, Prof No 39;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LFTPE 8
DB 117 LFTPE 124

RESULT 13
US-08-971-158-1
Sequence 1, Application US/08971158
Patent No. 6010879
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goli, Surya K.
TITLE OF INVENTION: HUMAN MITOCHONTRIAL CHAPERONE PROTEIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/971,158
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/824,875
FILING DATE: March 26, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0245-1 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555

TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 217 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: COLNOT08
CLONE: 1844121
US-08-971-158-1

Query Match 67.4%, Score 31, DB 3, Length 217;
Best Local Similarity 85.7%, Prof No 55;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFTPE 8
DB 178 LFTPE 184

RESULT 14
US-08-971-158-3
Sequence 3, Application US/08971158
Patent No. 6010879
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goli, Surya K.
TITLE OF INVENTION: HUMAN MITOCHONTRIAL CHAPERONE PROTEIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/971,158
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/824,875
FILING DATE: March 26, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0245-1 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 217 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1855878
US-08-971-158-3

Query Match 67.4%, Score 31, DB 3, Length 217;
Best Local Similarity 85.7%, Prof No 55;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0.

QY 2 LFETPVE 8
| | | | |
Db 178 LFETPVE 184

RESULT 15

US-09-437-054A-12
; Sequence 12, Application US/09437054A
; Patent No. 6316698
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Kinney, Anthony J.
; TITLE OF INVENTION: Plant Alpha-Glucosidase II Homologs
; FILE REFERENCE: BB1273 US NA
; CURRENT APPLICATION NUMBER: US/09/437,054A
; PRIOR FILING DATE: 2001-05-14
; PRIOR APPLICATION NUMBER: 60/107,909
; PRIOR FILING DATE: 1998-No. 6316698ember-10
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 12
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Zea mays
US-09-437-054A-12

Query Match 65.2%; Score 30, DB 4, Length 127,
Best Local Similarity 62.5%; Fred. No. 47,
Matches 5; Conservative 1, Mismatches 2, Indels 0; Gaps 0.

QY 1 NLFETPVE 8
| | | | |
Db 55 NLFETPVE 62

Search completed: September 5, 2002, 15:27:35
Job time: 320 sec

GenCore version 4.5
Copyright (c) 1994 - 2000 Compugen Ltd.

OM Protein - protein search, using sw model

File: us-09-744-804-38, 15, 26, 40, Search time: 15.43 seconds
(without alignments)
6,310 Million cell updates/sec

Title: US-09-744-804-38
Perfect score: 46
Sequence: 1 NIPETVEA 9

Scoring table: BLOSUM62
Gap: 10 0, Gapext: 0 5

Searched: 747574 seqs, 11073796 residues 747574
Total number of hits satisfying chosen parameters.

Minimum DB seq length: 0
Maximum DB seq length: 20000000

Post-processing: Minimum Match: 0%
Maximum Match: 100%
Listing first 45 summaries

Database: A_Geneseq_032802:*

1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

Pred No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Query Match	Length	EB ID	Description
1	46	100.0	9	21	Lactadherin (BA-16)
2	46	100.0	217	16	HMFG 46 KDa Antigen
3	46	100.0	218	16	HMFG 46 KDa Antigen
4	46	100.0	387	21	Human Lactadherin
5	46	100.0	387	21	Human Lactadherin
6	38	82.6	26	22	ABCI protein exten
7	38	82.6	1099	22	Human ABC1 protein
8	38	82.6	1144	22	Human ABC1 protein
9	38	82.6	1144	22	Human ABC1 protein
10	38	82.6	1528	22	Human ABC1 mutant
11	38	82.6	2130	22	Human ABC1 mutant

12	38	82.6	2143	21	AA58108	Human ABC1 cholest
13	38	82.6	2201	21	AA579380	Human ATP binding
14	38	82.6	2201	21	AA579380	Human ATP binding
15	38	82.6	2201	21	AA579380	Human ATP binding
16	38	82.6	2259	21	AA58107	Human ABC1 PHA-3 m
17	38	82.6	2260	21	AA58106	Human ABC1 cholest
18	38	82.6	2261	21	AA58092	Human ABC1 cholest
19	38	82.6	2261	21	AA58104	Human ABC1 cholest
20	38	82.6	2261	21	AA58105	Human ABC1 cholest
21	38	82.6	2261	21	AA58109	Human ABC1 cholest
22	38	82.6	2261	21	AA58110	Human ABC1 cholest
23	38	82.6	2261	21	AA58111	Human ABC1 cholest
24	38	82.6	2261	21	AA58112	Human ABC1 cholest
25	38	82.6	2261	21	AA58113	Human ABC1 cholest
26	38	82.6	2261	21	AA58114	Human ABC1 cholest
27	38	82.6	2261	21	AA58115	Human ABC1 cholest
28	38	82.6	2261	21	AA58116	Human ABC1 cholest
29	38	82.6	2261	21	AA58117	Human ABC1 cholest
30	38	82.6	2261	21	AA58118	Human ATP binding
31	38	82.6	2261	21	AA58028	Human ATP binding
32	38	82.6	2261	21	AA58029	Human protein seq
33	38	82.6	2261	21	AA58029	Human ABC1 mutant
34	38	82.6	2261	21	AA58029	Human ABC1 mutant
35	38	82.6	2261	21	AA58029	Human ABC1 mutant
36	38	82.6	2261	21	AA58029	Human ABC1 mutant
37	38	82.6	2261	21	AA58029	Human ABC1 mutant
38	38	82.6	2261	21	AA58029	Human ABC1 mutant
39	38	82.6	2261	21	AA58029	Human ABC1 mutant
40	38	82.6	2261	21	AA58029	Human ABC1 mutant
41	38	82.6	2261	21	AA58029	Human ABC1 mutant
42	38	82.6	2261	21	AA58029	Human ABC1 mutant
43	38	82.6	2261	21	AA58029	Human ABC1 mutant
44	38	82.6	2261	21	AA58029	Human ABC1 mutant
45	38	82.6	2261	21	AA58029	Human ABC1 mutant

ALIGNMENTS

RESULT 1

AA582843

ID: AA582843 standard; peptide; 9 AA.

AC: AA582843;

XX: 19 JUN 2003 (first entry)

XX: Lactadherin (BA-46) peptide fragment (tumour associated antigen).

FE: Tumour associated antigen peptide; TAA, cancer, carcinoma;

KW: treatment; prevention; cure; anti-tumour vaccine; metastases;

SW: breast, bladder, prostate, pancreas, ovary, thyroid, colon;

KW: stomach; carcinoma; MHC Class I; HLA A2; human;

KW: Major histocompatibility complex; uroplakin;

KW: prostate specific antigen; prostate specific membrane antigen;

KW: prostate acid phosphatase; mucin; lactadherin;

XX: teratocarcinoma derived growth factor; PSA; PSMA; PAP; CRP10-1.

XX: Homo sapiens.

OS: Homo sapiens.

XX: WO200005723-A1.

XX: 10 FEB 2000.

XX: 29-JUL-1999; 99WO-IL00417.

XX: 30-JUL-1998; 98IL-0135698.

XX: (YEDA) YEDA RES & DEV CO LTD.

XX: (HOF) BIO-TECHNOLOGY GEN CORP.

XX: Eisenbach L, Carmen L, Tirosh B, Bar haim E, Paz A, Fridkin M;

PI: Filizer-atlas C;

XX WPI: 2000-205463/18.
 DR Tumor associated antigen peptides, especially derived from uroplakin,
 XX useful as vaccines to prevent or cure cancers including breast,
 PT bladder, prostate, pancreas, ovary, thyroid, colon and stomach.
 PT Claim 17; Page 100; 11pp; English.
 XX Tumor associated antigen peptides (TAA) may be used for the
 CC treatment, prevention and cure of cancer or cancer metastases. The
 CC cancer may be breast, bladder, prostate, pancreas, ovary, thyroid,
 CC colon, stomach, head or neck cancer or a carcinoma. The tumor
 CC associated antigens are presentable to the immune system by HLA-A2
 CC molecules and are generally between 8 to 10 amino acids in length.
 CC The amino acids located at positions 2 and 9 of the tumor associated
 CC antigens are the anchor residues which participate in the binding to
 CC MHC class I molecules, more specifically HLA-A2. More tumor
 CC associated antigens are described in GENESQ records AAY82805-Y82882.
 CC Those tumor associated antigens described in records AAY82806-Y82834
 CC and AAY82855-Y82869 are derived from uroplakin, such as uroplakin II,
 CC uroplakin Ia, uroplakin III and uroplakin IV. Those described in
 CC records AAY82825-Y82829 are derived from prostate specific antigen
 CC (PSA). Those described in records AAY82830-Y82835 are derived from
 CC prostate specific membrane antigen (PSMA). Those described in
 CC records Y82836, AAY82839 are derived from prostate acid phosphatase
 CC (PAP). Those described in records AAY82840-Y82846 are derived
 CC from lactadherin (BA-46). Those described in records AAY82847-Y82854
 CC are derived from Murin and those described in records AAY82871-Y82882
 CC are derived from Teratocarcinoma derived growth factor (CRPTO-1).
 XX Sequence 9 AA;
 SQ

Query Match 100.0%; Score: 45; E-Value: 1.0e-11; Length: 9;
 Best Local Similarity 100.0%; Pos. Neg. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPVEA 9
 DB 1 nlfetpvea 9
 RESULT 2
 ID AAR77253
 AC AAR77253 standard; Protein: 217 AA.
 XX AAR77253;
 XX AAR77253;
 DT 21-NOV-1995 (first entry)
 DE HMF3 46 kDa antigen C-terminal region.
 KW HMF3, human milk fat globule, antigen; immunogen; vaccine; virucide;
 KW epithelium, tumor, breast cancer; monoclonal antibody; MAB.
 OS Homo sapiens.
 XX W09515171-A.
 PN 08-JUN-1995.
 XX 05-DEC-1994; 94WO-0513967.
 XX 03-DEC-1994; 94US-0162402
 XX (CANC-) CANCER RES FUND CONTRA COSTA.
 XX Ceriani RL, Larocca DJ, Peterson JA;
 WPI: 1995-215151/28
 XX 46 kD apparent molecular weight human milk fat globule antigen -
 PT used in assays to determine the presence of a cancerous tumour of
 epithelial origin, and in a vaccine against neoplastic tumours

PT used in assays to determine the presence of a cancerous tumour of
 PT epithelial origin, and in a vaccine against neoplastic tumours
 XX Claim 7; Page 41; 68pp; English.
 PS A partial cDNA clone hA45.1 (AA011155) for the 45 kDa HMF3 antigen,
 CC a major component of the apical surface of the normal breast
 CC epithelial cell, was obtained by PCR of cDNA from a lactating
 CC breast cDNA library. The C-terminal region of the encoded protein
 CC showed 43% identity to corresp. regions of human Factor-V and 38%
 CC to Factor-VIII
 XX Sequence 217 AA;
 SQ

Query Match 100.0%; Score: 46; E-Value: 1.0e-11; Length: 217;
 Best Local Similarity 100.0%; Pos. Neg. 9.1e+19;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPVEA 9
 DB 24 nlfetpvea 32
 RESULT 3
 ID AAR77254 standard; Protein: 218 AA.
 XX AAR77254;
 XX AAR77254;
 DT 21-NOV-1995 (first entry)
 DE HMF3 46 kDa antigen partial sequence.
 KW HMF3, human milk fat globule, antigen; immunogen; vaccine; virucide;
 KW epithelium; tumor; breast cancer; monoclonal antibody; MAB.
 OS Homo sapiens.
 XX Key Location/Qualifiers
 FT Modified-site 69..71
 FT /label= N-glycosylation_site
 FT Modified-site 156..158
 FT /label= N glycosylation_site
 FT Modified-site 160..162
 FT /label= N-glycosylation_site
 FT Modified-site 181..183
 FT /label= N-glycosylation_site
 XX W09515171-A.
 PN 08-JUN-1995.
 XX 05-DEC-1994; 94WO-0513967.
 XX 03-DEC-1994; 94US-0162402.
 XX (CANC-) CANCER RES FUND CONTRA COSTA.
 XX Ceriani RL, Larocca DJ, Peterson JA;
 WPI: 1995-215151/28.
 XX N-PSDB; AAQ91199.
 XX 46 kD apparent molecular weight human milk fat globule antigen -
 PT used in assays to determine the presence of a cancerous tumour of
 PT epithelial origin, and in a vaccine against neoplastic tumours
 XX Example 7; Page 38-40; 68pp; English.
 PS A partial cDNA clone hA45.1 (AA011155) for the 45 kDa HMF3 antigen,
 CC a major component of the apical surface of the normal breast
 CC epithelial cell, was obtained by PCR of cDNA from a lactating

CC breast cDNA library.
XX Sequence 218 AA;
SQ

Query Match 100.0%; Score 46; DB 16; Length 218;

Best Local Similarity 100.0%; Pred. No. 0.45;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPVEA 9
XXXXXXXXXX
Db 25 nlfetpvea 33

RESULT 4

AAR77252

ID AAR77252 standard; Protein; 387 AA.

XX AC AAR77252;

XX 21-NOV-1995 (first entry)

XX HMFG 46 kDa antigen.

XX HMFG; human milk fat globule; antigen; immunogen; vaccine; virucide;
XX epithelium; tumor; breast cancer; monoclonal antibody; Mab.

XX Homo sapiens.

XX W09515171-A.

XX 08-JUN-1995.

XX 05-DEC-1994; 94W0-0513667.

XX 03-DEC-1993; 930S-0152402.

XX (CANC-) CANCER PES FUND CONTEA COSTA.

XX Ceriani RL, Larocca DJ, Peterson JA;

XX WPI; 1945 015151/28.

XX N-PSDB; AAG01108.

XX 46 kD apparent molecular weight human milk fat globule antigen -
XX used in assays to determine the presence of a carcinosarcoma of
XX epithelial origin, and in a vaccine against neoplastic tumours

XX Claim 6; Page 46-47, 68pp, English

XX A complete cDNA sequence for the 46 kDa HMFG antigen, a major
XX component of the apical surface of the normal breast epithelial
XX cell, was obtained by PCR and RACE methods. cDNA clones can be
XX used to prepare MAbs for use in immunotherapy, immunohistochemistry,
XX prognosis, diagnosis and therapy. Recombinant antigen can
XX be expressed in prokaryotic or (glycosylated) in eukaryotic cells.

XX Sequence 387 AA;

Query Match 100.0%; Score 46; DB 16; Length 387;

Best Local Similarity 100.0%; Pred. No. 0.45;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPVEA 9
XXXXXXXXXX
Db 194 nlfetpvea 202

RESULT 5

AA94453

ID AA94453 standard; Protein; 387 AA.

XX

AC AA94453;

XX 11-SEP-2000 (first entry)

XX Human lactadherin protein.

XX Human lactadherin, MGF-F8, anti tumor, immune response;
XX exosome; dendritic cell.

XX Homo sapiens.

XX Key

XX Peptide

XX Location/Qualifiers

XX 1..23

XX /label= Secretion_signal

XX 24..387

XX /label= Lactadherin

XX 46..48

XX /label= Integrin_binding_site

XX EP1004664-A1.

XX 31-MAY-2000.

XX 24-NOV-1998; 98EP-0402925.

XX 24-NOV-1998; 98EP-0402925.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX (CURT-) INST CURTE.

XX WPI; 2000-352547/31.

XX N-PSDB; AAA27140.

XX Chimeric isolated (human) lactadherin polypeptide that functions as an

XX adaptor of cross priming to eliminate pathogenic antigens, e.g. in

XX tumors -

XX Example 3; Page 12; 20pp; English.

XX Lactadherin protein was found in exosomes produced by dendritic cells.

XX The protein is involved in the phagocytosis of particulate antigens by

XX dendritic cells. Exosomes produced by dendritic cells exposed to

XX tumor antigens induce potent immune responses. Lactadherin or variants

XX of it may be used in the mediation of an immune response. Variants of

XX lactadherin may be used for inhibition and/or stimulation of the

XX cross priming of antigens and stimulation of the phagocytosis of

XX antigens by dendritic cells. Compositions derived from lactadherin can

XX also be used to monitor an immune response, more specifically a CD1

XX (cytotoxic lymphocyte) response and also to produce CD1 specific

XX for a selected antigen. The present sequence is the human

XX lactadherin protein.

XX Sequence 387 AA;

QY

1 NLFETPVEA 9

XXXXXXXXXX

Db 194 nlfetpvea 202

RESULT 6

AAB62694

ID AAB62694 standard; Peptide; 26 AA.

XX AAB62694;

XX 06-AUG-2001 (first entry)

XX ABC1 protein external domain TM5-TM6 fragment (residues 795-820).

DE

Query Match 100.0%; Score 46; DB 21; Length 397;

Best Local Similarity 100.0%; Pred. No. 0.35;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

XX ABC1, antilipemic; cholesterol; inhibitor; low density lipoprotein; LDL.
XX Homo sapiens.
XX W0260132184-A2.
XX 10-MAY-2001.
XX 01-NOV-2000; 2000WO-US30109
XX 01-NOV-1999; 990S-0162803.
XX 30-JUN-2000; 2000US-0215564.
XX (WISC ) WISCONSIN ALUMNI RES FOUND.
XX Attie AB, Cook M, Gray-Keller MP, Hayden MK, Pinstone S;
XX Brooks-Wilson A;
XX WPI; 2001-335779/35
XX New method for inhibiting cholesterol uptake in the gut comprises
XX administration of an inhibitor of an ABC1 protein
XX Disclosure; Page 9, 41pp; English.
XX The invention relates to a new method for inhibiting cholesterol uptake
XX in the gut that comprises administration of an inhibitor of an ABC1
XX protein. The method is useful for: lowering levels of LDL (low density
XX lipoprotein) cholesterol by reducing the activity of ABC1 protein in the
XX intestinal cells and the abundance of the ABC1 protein in the individual
XX by inhibiting the activity of the protein; identifying drugs that can
XX lower serum cholesterol and LDL levels comprising assaying the drug to
XX test if it can bind to an ABC1 protein; testing LDL cholesterol lowering
XX agents; and for modulation of ABC1 biological activity. Sequences
XX AAB62691-97 represent predicted external domain of ABC1 protein.
XX Sequence 26 AA;
XX
XX Query Match 82.5%; Score 38; DB 22; Length 26;
XX Best Local Similarity 87.5%; Pred. No. 0.82;
XX Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0.
XX
XX QY 1 NLFETPVE 8
XX IIII:III
XX Db 14 nlfespe 21
XX
XX RESULT 7
XX AAB62691
XX ID AAB62691 standard; Protein; 1089 AA.
XX AC AAB62691;
XX XX
XX DT 06-AUG-2001 (first entry)
XX DE Human ABC1 protein partial fragment
XX KW ABC1, antilipemic, cholesterol, inhibitor, low density lipoprotein; LDL.
XX OS Homo sapiens.
XX XX
XX PN W0260132184-A2
XX XX
XX DT 10-MAY-2001
XX XX
XX PF 01-NOV-2000; 2000WO-US30109.
XX XX
XX PR 01-NOV-1999; 990S-0162803.
XX FF 30-JUN-2000; 2000US-0215564
XX XX
XX PA (WISC ) WISCONSIN ALUMNI RES FOUND.

```

```

XX Attie AB, Cook M, Gray-Keller MP, Hayden MK, Pinstone S;
XX Brooks-Wilson A;
XX WPI; 2001 335779/35.
XX N F3DB; AAF83826.
XX New method for inhibiting cholesterol uptake in the gut comprises
XX administration of an inhibitor of an ABC1 protein
XX Disclosure; Page 37-; 41pp; English.
XX The invention relates to a new method for inhibiting cholesterol uptake
XX in the gut that comprises administration of an inhibitor of an ABC1
XX protein. The method is useful for: lowering levels of LDL (low density
XX lipoprotein) cholesterol by reducing the activity of ABC1 protein in the
XX intestinal cells and the abundance of the ABC1 protein in the individual
XX by inhibiting the activity of the protein; identifying drugs that can
XX lower serum cholesterol and LDL levels comprising assaying the drug to
XX test if it can bind to an ABC1 protein; testing LDL cholesterol lowering
XX agents; and for modulation of ABC1 biological activity. The present
XX sequence represents a human ABC1 protein partial fragment.
XX Note: Pages containing part of the present sequence seems to be missing
XX from the specification; the protein seems to contain at least 2261 amino
XX acid residues as found by translating the corresponding DNA sequence.
XX Sequence 1089 AA,
XX
XX Query Match 82.5%; Score 38; DB 22; Length 1089
XX Best Local Similarity 87.5%; Pred. No. 43;
XX Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 NLFETPVE 8
XX IIII:III
XX Db 808 nlfespe 815
XX
XX RESULT 8
XX AA002184
XX ID AA002184 standard; Protein; 1144 AA.
XX AC AA002184,
XX XX
XX DT 12-SEP-2001 (first entry)
XX DE Human ABC1 mutant polypeptide #8.
XX KW Human; ABC1 gene; atherosclerosis, reverse transport, cholesterol;
XX cardiovascular; neurological; Tangier disease; LCAT deficiency; mutant;
XX lecithin:cholesterol acetyltransferase; malaria; diabetes; mutain.
XX OS Homo sapiens.
XX XX
XX FH Key location/Qualifiers
XX FT Misc-difference 1114 /note- "Frameshift at amino acid 1114 introduces a stop
XX FT codon at amino acid 1144"
XX XX
XX FN W0260130848-A2.
XX XX
XX PD 03-MAY-2001.
XX XX
XX PF 26-OCT-2000; 2000WO-EP10886.
XX XX
XX PP 26-OCT-1999; 99FP-0402668.
XX PR 01-MAR-2000; 2000US-0186280.
XX XX
XX PA (AVET ) AVENTIS PHARMA SA.
XX XX
XX PI Lemelle P, Rosier-Medras M, Arnould-Pepinque I, Prajes C, Naudin L;
XX P1 Lemelle C, Buverger N, Jaye M, Scarfoss CH, Remaley A, Brewer BB;
XX Dean M;

```

```

XX WPI: 2001-316327/23.
DR N-PSDB: AAS06133.
XX
XX New human ABC1 nucleic acids and polypeptides for treating
PT atherosclerosis, malaria and diabetes.
XX
XX Claim 46: Page 287-292; 368pp; English.
PS
CC The sequence represents the amino acid sequence of human ABC1 mutant
CC polypeptide #8. The polypeptide and nucleic acid sequence, primers and
CC probes derived from the ABC1 sequence, and vectors are useful for the
CC prevention of atherosclerosis, in a subject affected by a dysfunction in
CC the reverse transport of cholesterol. The polypeptide encoded by the ABC1
CC gene is useful for screening for an active ingredient for the prevention
CC or treatment of a disease resulting from dysfunction in the reverse
CC transport of cholesterol. The nucleic acids and polypeptides are also
CC useful for treating and preventing cardiovascular and neurological
CC pathologies, and other diseases e.g. Tangier disease, lecithin-
CC cholesterol (LCAT) deficiency, malaria and diabetes.
XX
SQ Sequence 1144 AA:

Query Match      82.6%; Score 38; DB 22; Length 1144;
Best Local Similarity 87.5%; Pred. No. 45;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPVE 8
DB 808 nlfespve 815
      |||||
      |||||

RESULT 10
AAU02187
ID AAU02187 standard; Protein: 1525 AA.
XX
AC AAU02187;
XX
DT 12-SEP-2001 (first entry)
DE Human ABC1 mutant polypeptide #11.
XX
KW Human; ABC1 gene; atherosclerosis; reverse transport; cholesterol;
KW cardiovascular; neurological; Tangier disease; LCAT deficiency; mutant;
KW lecithin-cholesterol acetyltransferase; malaria; diabetes; nuclein.
XX
OS Homo sapiens.
XX
FH Location/Qualifiers
FT Misc-difference 1525
FT /note= "Wild type Trp substituted by Stop codon"
XX
PN W0200130848-A2.
XX
PD 03-MAY-2001.
XX
PF 26-OCT-2000; 2000WO-EP10886.
XX
PR 26-OCT-1999; 99EP-0402668.
PP 01-MAR-2000; 2000US-0186260.
XX
PA (AVET ) AVENTIS PHARMA SA.
XX
PI Denelle P, Rosier-Montus M, Arnould-Requique J, Prades C, Naudin L;
PI Lemoine C, Duvrquier N, Jaye M, Searfoss GH, Komaley A, Brewer HB;
PI Beau M;
XX
DR WPI: 2001-316327/23.
DR N-PSDB: AAS06133.
XX
XX New human ABC1 nucleic acids and polypeptides for treating
PT atherosclerosis, malaria and diabetes.
XX
XX Claim 46: Page 305-312; 368pp; English.
PS
CC The sequence represents the amino acid sequence of human ABC1 mutant
CC polypeptide #11. The polypeptide and nucleic acid sequence, primers and

```

CC probes derived from the ABC1 sequence, and vectors are useful for the
 CC prevention of atherosclerosis, in a subject affected by a dysfunction in
 CC the reverse transport of cholesterol. The polypeptide encoded by the ABC1
 CC gene is useful for screening for an active ingredient for the prevention
 CC or treatment of a disease resulting from dysfunction in the reverse
 CC transport of cholesterol. The nucleic acids and polypeptides are also
 CC useful for treating and preventing cardiovascular and neurological
 CC pathologies, and other diseases e.g. Tangier disease, lecithin-
 CC cholesterol (LCAI) deficiency, malaria and diabetes.

XX Sequence 1525 AA;

Query Match 92.6%; Score 38; DB 22; Length 1525;
 Best Local Similarity 87.5%; Pred. No. 62;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QV 1 NLFETPVE 8
 |||||
 Db 808 nlfespe 815

RESULT 11

AAU02190
 ID AAU02190 standard; Protein: 2130 AA.

XX AC
 AC AAU02190;

DT 12 SEP 2001 (first entry)

DE Human ABC1 mutant polypeptide #14.

XX Human, ABC1 gene, atherosclerosis; reverse transport; cholesterol;
 KW cardiovascular, neurological, Tangier disease, LCAT deficiency, mutant,
 KW lecithin cholesterol, acetyltransferrase, malaria, diabetes, mutin.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Misc-difference 2132

FT /note "Frameshift at amino acid 2112 introduces a stop
 codon at amino acid 2130"

XX WC:0013044P A2.

XX XX 03-MAY-2001

XX PR 26-OCT-2000; 2000G0-IR00592

XX PR 26-OCT-1999; 99EP-0402668.

XX PR 01-MAR-2000; 2000US-0186260.

XX PA (AVET) AVENTIS PHARMA SA.

XX PI Denelle P, Rosier-Montus M, Arnould-Requigne I, Prades C, Naudin I;

XX PI Lemoine C, Duverger N, Baye M, Searfoss GH, Pomaleu A, Prewer HR;

XX PI Dean M;

XX DR WPI: 2001-016327/33.

XX DR N-PSDB: AAC06138.

XX New human ABC1 nucleic acids and polypeptides for treating
 PT atherosclerosis, malaria and diabetes -

XX PS Claim 46; Page 335 346, 368ff; English.

XX The sequence represents the amino acid sequence of human ABC1 mutant
 CC polypeptide #14. The polypeptide and nucleic acid sequence, primers and
 CC probes derived from the ABC1 sequence, and vectors are useful for the
 CC prevention of atherosclerosis, in a subject affected by a dysfunction in
 CC the reverse transport of cholesterol. The polypeptide encoded by the ABC1
 CC gene is useful for screening for an active ingredient for the prevention
 CC or treatment of a disease resulting from dysfunction in the reverse

CC transport of cholesterol. The nucleic acids and polypeptides are also
 CC useful for treating and preventing cardiovascular and neurological
 CC pathologies, and other diseases e.g. Tangier disease, lecithin-
 CC cholesterol (LCAI) deficiency, malaria and diabetes.

XX Sequence 2130 AA;

Query Match 83.6%; Score 39; DB 22; Length 2130;
 Best Local Similarity 87.5%; Pred. No. 88;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QV 1 NLFETPVE 8
 |||||
 Db 808 nlfespe 815

RESULT 12

AAAB38108

IC AAB38108 standard; Protein: 2143 AA.

XX XX

AC AAB38108;

XX DT 29-JAN-2001 (first entry)

DE Human ABC1 cholesterol transporter PHA-1 mutant protein (R2144STOP).

XX Human ABC1 cholesterol transporter, chromosome 9q31;

KW ATP binding cassette, ABC deficiency disorder, high density lipoprotein;

KW Tangier disease; ID: familial HDL deficiency; PHA; polymorphism;

KW cardiovascular disease, coronary artery disease, coronary restenosis;

KW cerebrovascular disease; peripheral vascular disease;

KW Alzheimer's disease; Niemann-Pick disease; Huntington's disease;

KW X-linked adrenoleukodystrophy; cancer; gene therapy; genetic diagnosis;

KW prognosis, prophylaxis, drug screening, transgenic animal; mutant;

XX KW

OS Homo sapiens.

XX WC:00055318 A2.

XX DT 21-SEP-2000.

XX PR 15-MAR-2000; 2000G0-IR00592.

XX PR 15-MAR-1999; 99US-0124702.

XX PR 08-JUN-1999; 99US-0138048.

XX PR 17-JUN-1999; 99US-0149600.

XX PR 01-SEP-1999; 99US-0151977.

XX PA (OVRP-) UNIV BRITISH COLUMBIA

XX PA (XENO-) XENON BIORESEARCH INC.

XX PI Hayden MR, Wilson AR, Pimstone SN.

XX DP WPI: 2000-58759/55

XX DR N-PSDB: AAC69389.

XX New ABC1 polypeptide is useful for treating diseases associated with
 PT ABC1 biological activity, e.g. Alzheimer's disease, Huntington's
 PT disease and cancer -

XX Examples; Page -: 229pp; English.

XX The invention relates to the human ABC1 cholesterol transporter protein
 CC (B38082) and to nucleic acid sequences (569120) which encode it. ABC1 is
 CC a member of the ATP-binding cassette (ABC transporter) superfamily of
 CC proteins, and plays a crucial role in cholesterol transport, particularly
 CC intracellular cholesterol trafficking in monocytes and fibroblasts, being
 CC involved in cholesterol efflux from the cell. The gene encoding ABC1 is
 CC located on chromosome 9q31, and mutations in this gene are associated
 CC with two genetic HDL (high density lipoprotein) deficiency disorders,
 CC Tangier disease (TD) and familial HDL deficiency (FH). These diseases

are distinguishable in that TH is an autosomal recessive disorder, while
 FHA is inherited as an autosomal dominant trait; low levels of HDL ("good
 cholesterol") in the blood correlate with a high risk of cardiovascular
 disease, particularly coronary artery disease, but also cerebrovascular
 disease, coronary restenosis, and peripheral vascular disease.
 Conversely, a high level of HDL has protective effects against
 cardiovascular disease. The invention provides genetic constructs and
 transgenic cells and non-human animals comprising human ABCA1 nucleic
 acids, and methods of gene therapy for the treatment or prevention of
 cardiovascular disease comprising the administration of an expression
 vector encoding ABCA1 or an active fragment thereof. The invention also
 encompasses compounds which mimic ABCA1 activity, compounds which
 stimulate ABCA1 expression and methods of screening for such compounds.
 It further relates to methods for determining whether a patient has an
 increased risk for cardiovascular disease due to polymorphisms in the
 ABCA1 gene. Human ABCA1 proteins and nucleotides can be used to treat
 or prevent cardiovascular disease, especially coronary artery disease,
 cerebrovascular disease, coronary restenosis or peripheral vascular
 disease. They may also be used in the treatment of diseases associated
 with ABCA1 biological activity, such as Alzheimer's disease, Niemann-Pick
 disease, Huntington's disease, X-linked adrenoleukodystrophy and cancer.
 The invention specifically excludes proteins with the exact amino acid
 sequences of GenBank Accession No. U04100.1 and X75426, and the nucleic
 acid with the exact sequence as GenBank Accession No. AF012376.1. The
 present sequence represents a mutant human ABCA1 cholesterol transporter
 associated with an altered cholesterol level and therefore an altered
 risk of cardiovascular disease.
 Note: The present sequence is not shown in the specification, but is
 derived from the native human ABCA1 shown on pages 152-157

Sequence 2143 AA;

Query Match 82.6%; Score 30; DB 21; Length 2143;
 Best Local Similarity 87.5%; Pred. No. 88;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPVE 8
 |||||
 Db 808 nlfespe 815

RESULT 13
 AAY79380
 ID AAY79380 standard; Protein, 2201 AA.
 AC AAY79380;
 XX
 XX
 DT 01-AUG-2000 (first entry)
 XX
 DE Human ATP binding cassette ABCA1 (ABCA1) protein.

ABCA1: ABCA1: ATP binding cassette, human, cholesterol,
 interleukin-1 beta, transporter, inflammation; septic shock;
 rheumatoid arthritis; Tangier disease; hypertriglyceridemia;
 splenomegaly; atherosclerosis; lipid disorder; dyslipidemia;
 psoriasis; lupus erythematosus; diagnosis; therapy.

Homo sapiens.

WC200018912-A2.

06-APR-2000.

21-SEP-1999; 99WO-EP06991.

25-SEP-1998; 98US-0101706.

(FARB) BAYER AG.

Schmitz G, Klucken J;

WPI: 2000-293151/25.

DE N-PSDB, AA294734.

Adenosine triphosphate binding proteins useful for identifying agents
 for treating atherosclerosis and other inflammatory disorders -
 Claim 6, Page 94 105, 154pp; English.

The present sequence is that of human ATP binding cassette protein
 ABCA1 (ABCA1), the human homologue of mouse ABCA1 (94% identity).
 The sequence was deduced from ABCA1 cDNA (see AA294734); identified in
 a differential screening of human cDNAs for cholesterol sensitive
 genes, the ABCA1 gene maps to chromosome 9q22-q31. Dysregulated
 ABCA1 is the gene locus involved in the HDL deficiency syndrome
 Tangier disease, associated with hypertriglyceridemia and
 splenomegaly. ABCA1 is also a transporter for interleukin-1 beta,
 making the gene a candidate for treatment of inflammatory diseases
 such as rheumatoid arthritis and septic shock. The invention
 also provides other cholesterol-sensitive ABC genes (see AA294735-63)
 that can be used for diagnostic and therapeutic applications,
 and for biochemical or cell-based assays to screen for
 pharmacologically active compounds useful for the treatment of
 lipid disorders, atherosclerosis or other inflammatory diseases
 such as psoriasis and lupus erythematosus.

Sequence 2201 AA;

Query Match 82.5%; Score 38; DB 21; Length 2201;
 Best Local Similarity 87.5%; Pred. No. 91;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPVE 8
 |||||

Db 748 nlfespe 755

RESULT 14

AAE13021

ID AAE13021 standard; Protein, 2201 AA.

AC AAE13021;

28 JAN 2002 (first entry)

Human ATP binding cassette transporter 1 (ABCA1) protein.

Human, ATP binding cassette transporter 1; ABCA1; coronary heart disease;
 dermatological; atherosclerosis; cardiovascular; inflammatory disease;
 psoriasis; lipid disorder; antibacterial; septic shock; gene therapy;
 immunosuppressive; lupus erythematosus; rheumatoid arthritis.

Homo sapiens.

key location/qualifiers

Misc-difference 978 /note "Encoded by CAC"

Misc-difference 980 /note= "Encoded by AAC"

EP1136552-A1.

26-SEP-2001.

29-MAR-2000; 2000EP-0105820

20-MAR-2000; 2000EP-0105820.

(FARB) BAYER AG.

Schmitz G, Bodzioch M;

WPI: 2001-640388/74.

N-PSDB, AA221325.

XX New adenosine triphosphate binding cassette transporter-1 gene
 PT polymorphisms, useful for diagnosing and treating lipid disorders,
 PT cardiovascular diseases and inflammatory diseases -
 XX
 PS Claim 2, Page 23-25, 48pp, English.
 XX
 CC The invention relates to four common polymorphisms in the gene encoding
 CC ATP binding cassette transporter-1 (ABCL). ABCL is associated with
 CC decreased ApoA-I mediated efflux of cholesterol. The polymorphisms in
 CC ABCL directly affects cellular lipid homeostasis, which is a key factor
 CC in the atherogenic processes. The ABCL polymorphisms are useful for
 CC diagnosing and treating lipid disorders, cardiovascular diseases
 CC (coronary heart disease, atherosclerosis) and inflammatory diseases
 CC (psoriasis, lupus erythematosus). The identification of ABCL as a
 CC transporter for interleukin-beta (IL-1beta) identifies this gene as
 CC a candidate for treatment of inflammatory diseases including rheumatoid
 CC arthritis and septic shock. The present sequence is human ABCL protein
 XX
 SQ Sequence 2201 AA;

Query Match 82.5%; Score 38; DB 22; Length 2201;
 Best Local Similarity 87.5%; Prod No 91;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLPETPVE 8
 |||||
 Db 748 nlfespe 755

RESULT 15

AAM50227

ID AAM50227 standard; Protein; 2201 AA.

XX AC AAM50227;

XX DT 07-JAN-2002 (first entry)

XX DE Human ATP binding cassette transporter 1 (ABCL).

XX KW ATP binding cassette transporter 1, ABCL, human, lipid disorder,
 KW cholesterol, cardiovascular disease, inflammatory disease,
 KW antiinflammatory; antilipemic; antipsoriasis; dermatological;
 KW tangier disease; coronary heart disease; diagnosis; gene therapy.

XX OS Homo sapiens.

XX XX EPI136554-A1.

XX XX 26-SEP-2001.

XX XX 24-MAR-2000; 2000EP-0106401.

XX XX 24-MAR-2000; 2000EP-0106401.

XX XX (FARB) BAYER AG.

XX XX Schmitz G, Bodzioch M;

XX XX WPI: 2001-640389/74.

XX XX N-PSDB; AAI70314.

XX New adenosine triphosphate binding cassette transporter gene
 PT polymorphisms, useful for diagnosing and treating lipid disorders,
 PT cardiovascular diseases and inflammatory diseases -

PS Example 1; Page 19-25; 41pp; English.

XX The present sequence is that of the human adenosine triphosphate
 CC (ATP) binding cassette transporter 1 (ABCL) protein. The invention
 CC provides 4 common polymorphisms in the ABCL gene (see AAI70314),
 CC all of which alter the amino acid sequence of ABCL and therefore

CC may affect its function. The 2 most common polymorphisms are both
 CC associated with a decreased in vitro ApoA-I mediated efflux of
 CC cholesterol from monocyte-derived phagocytes, a feature typical of
 CC tangier disease. 3 Of the variants are significantly increased in
 CC a population of men having low high density lipoprotein-cholesterol
 CC levels and established coronary heart disease (CHD) relative to
 CC CHD-free control subjects. The use of the provided ABCL polymorphisms
 CC for the diagnosis and treatment of lipid disorders, cardiovascular
 CC diseases, and inflammatory diseases (e.g., psoriasis, lupus
 CC erythematosus) is claimed. Modulation of ABCL transcripts or
 CC proteins by antisense or ribzyme technology or RNA decoys is also
 CC claimed.

XX SQ Sequence 2201 AA,

Query Match 82.5%; Score 38; DB 22; Length 2201;
 Best Local Similarity 87.5%; Prod No 91;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLPETPVE 8

|||||

Db 748 nlfespe 755

Search completed: September 5, 2002, 15:26:31
 Job time: 371 sec

GenScore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using a model
Run on: September 5, 2002, 15:29:51 : Search time 49.78 seconds
(without alignments)
12,393 Million cell updates/sec

Title: US-09-744-804-38
Perfect score: 46
Sequence: 1 NLFETPVEA 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match: 0%
Maximum Match: 100%
Listing first 45 summaries

Database: PIR71:
1: pir1:
2: pir2:
3: pir3:
4: pir4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	100.0	218	2 A47285	milk fat globule p
2	39	82.6	2061	2 A54774	ATP binding cass
3	37	80.4	381	2 AC2631	conserved hypoth
4	37	80.4	461	2 G97413	hypothetical prote
5	36	78.3	401	2 S65138	glycoprotein antiq
6	36	78.3	427	2 S74211	PAS-6/7 protein pr
7	36	78.3	655	2 A10910	probable membrane
8	35	76.1	566	2 G60588	poly A polymerase
9	35	76.1	658	2 T40107	hypothetical 57.4
10	34	73.9	177	2 T08448	hypothetical prote
11	34	73.9	241	2 S73277	beta-lactamase
12	34	73.9	409	2 T11743	ph47 protein - plq
13	34	73.9	744	2 T27674	hypothetical prote
14	34	73.9	762	2 T63165	hypothetical prote
15	33	71.7	249	2 P91282	probable periplasm
16	33	71.7	294	2 S94649	beta-lactamase (EC
17	33	71.7	343	2 T26784	hypothetical prote
18	33	71.7	405	2 G84280	cytochrome P450 11
19	33	71.7	464	2 T16889	protein kinase (p
20	33	71.7	535	2 S76953	WAM dehydrogenase
21	33	71.7	612	2 T11830	methylnalonyl CoA
22	33	71.7	713	2 G98287	methylnalonyl CoA
23	33	71.7	713	2 AB2297	methylnalonyl CoA
24	33	71.7	714	2 S87543	hypothetical prote
25	32	69.6	179	2 H87418	hypothetical prote
26	32	69.6	196	2 D84074	hypothetical prote
27	32	69.6	213	2 A31498	SAM-dependent meth
28	32	69.6	291	2 S93230	hypothetical prote
29	32	69.6	312	2 A54555	beta-lactamase (EC

ALIGNMENTS

RESULT 1

A47285
milk fat globule protein human (fragment)
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 28 May-1999
C:Accession: A47285
Kolarow, P., Petersen, J.A., Orlean, R., Kufiyoshi, J., Bistrain, A.M., Ceriani, R.L.
Cancer Res. 51, 4994-4998, 1991
A:Title: A Mr 46,000 human milk fat globule protein that is highly expressed in huma
A:Reference number: A47285; M010:91371351
A:Accession: A47285
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-218 <EAS>
A:Cross references: CB:356151; NID:q275396; PID:AAH19771.1; PID:q275397
C:Superfamily: milk fat globule protein; discoidin 1 amino-terminal homology; EGF hom
F.1 567domain, discoidin 1 amino terminal homology (fragment) <DN1>
F.60-218/Domain, discoidin 1 amino terminal homology <DN2>

Query Match 100.0%; Score 46; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 0.048;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CT 1 NLFETPVEA 9
CL 25 NLFETPVEA 33

RESULT 2

A54774
ATP binding cassette transporter ABC1 mouse
C:Species: Mus musculus (house mouse)
C:Date: 27-Apr-1997 #sequence_revision 27-Apr-1997 #text_change 02 Feb-2001
C:Accession: A54774
Kuchel, M.F., Benfante, P., Parary, S., Martel, M.G., Gimble, G.
Genomics 21, 150-159, 1994
A:Title: Cloning of the novel ABC transporters mapping on human chromosome 9.
A:Reference number: A54774; M010:94375008
A:Accession: A54774
A:Molecule type: mRNA
A:Residues: 1-2201 <EAS>
A:Cross references: CB:356151; NID:q275396; PID:AAH19771.1; PID:q275397
C:Superfamily: conserved ATP binding cassette protein; ATP-binding cassette homology
C:Keywords: ATP, duplication, nucleotide binding; p-loop
F.856 1047/Domain, ATP binding cassette homology <ABC1>
F.973 863/Region, nucleotide-binding motif A (p-loop)
F.1869 2902/Domain, ATP binding cassette homology <ABC2>
F.1886 1893/Region, nucleotide-binding motif A (p-loop)

Query Match 92.6%; Score 39; DB 2; Length 2201;

Best Local Similarity 87.58; Pred. No. 32;
Matches 7; Conservative 1; Mismatches 6; Gaps 0.

QY 1 NLFETPVE 8
|||||
DB 748 NLFESPVE 755

RESULT 3
AC2631
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11 Jan 2002 #text_change 11-Jan-2002
C:Accession: AC2631
R:Wood, D.W.; Schubert, J.C.; Kaul, R.; Marks, D.; Chen, L.; Wood, G.F.; Chen, Y.; Wang, J.; Krup, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.E.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577, PMID:11743193
A:Accession: AC2631
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-361 <KOR>
A:Cross-references: GB:AF028688, FIDB:AA141465.1, FIDB:G17738790, GSPDB:G000186
A:Experimental source: strain C58 (bupont)
A:Genetics:
A:Gene: Atg0446
A:Map position: circular chromosome

Query Match 80.48; Score 37; DB 2; Length 381;
Best Local Similarity 77.88; Pred. No. 7;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0.

QY 1 NLFETPVEA 9
|||||
DB 180 NLFETPPEA 188

RESULT 4
D97413
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: D97413
R:Goodner, R.; Hickler, G.; Gattwood, S.; Miller, N.; Blanchard, M.; Gurelli, R.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Seery, C.; Lappas, C.; Markel, R.
Science 294, 2323-2329, 2001
A:Title: Genome Sequence of the Plant Pathogen and Bio-technology Agent Agrobacterium tumefaciens
A:Reference number: A97359; PMID:11743194
A:Accession: D97413
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-461 <KOR>
A:Cross-references: GB:AF028688; FIDB:AAK00261.1, FIDB:G1555606; GSPDB:G000186
A:Genetics:
A:Gene: AGR_C_790
A:Map position: circular chromosome

Query Match 80.48; Score 37; DB 2; Length 461;
Best Local Similarity 77.88; Pred. No. 8.9;
Matches 7; Conservative 1; Mismatches 0; Gaps 0.

QY 1 NLFETPVEA 9
|||||
DB 260 NLFETPPEA 268

RESULT 5

S65138
glycoprotein antigen M657/53, mammary gland - bovine (fragment)
N:Alternate names: glycoprotein component 15, major fat-globule membrane protein, M657-E
C:Species: Bos primigenius taurus (cattle)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998
C:Accession: S65138; G48394
R:Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.
Biochim. Biophys. Acta 1245, 385-391, 1995
A:Title: Molecular cloning of glycoprotein antigens M657/53 recognized by monoclonal
A:Reference number: S65138; MIMD:96125746
A:Accession: S65138
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-401 <KOR>
R:Mother, T.H.; Rungberg, L.P.; Lane, W.S.
Biochem. Mol. Biol. Int. 29, 545-554, 1993
A:Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-
A:Keywords: glycoprotein
A:Cross-references: NCBI:131457
A:Note: sequence extracted from NCBI bank-one (NCBI:131457)
C:Superfamily: milk fat globule protein, discoidin I amino-terminal homology; EGF hom
F:1-42/Domain: EGF homology (fragment) <EG>
F:40-79/Domain: EGF homology <EG>
F:82-133/Domain: discoidin I amino-terminal homology <DN>
F:243-401/Domain: discoidin I amino-terminal homology <DN>

Query Match 78.38; Score 36; DB 2; Length 401;
Best Local Similarity 75.08; Pred. No. 12;
Matches 6; Conservative 2; Mismatches 0; Gaps 0.

QY 1 NLFETPVE 8
|||||
DB 208 NLFETPPE 215

RESULT 6
S74211
C:Species: Bos primigenius taurus (cattle)
C:Date: 24-Jun-1997 #sequence_revision 12-Jun-1997 #text_change 20-Jun-2000
C:Accession: S74211; S78114; S24181
R:Huvaric-Juric, J.; Andersen, M.H.; Rungberg, L.; Rasmussen, J.T.; Petersen, T.E.
Eur. J. Biochem. 249, 628-636, 1996
A:Title: Characterization of glycoprotein PAS-6/7 from membranes of bovine milk fat
A:Reference number: S74211; MIMD:9708954
A:Accession: S74211
A:Molecule type: mRNA
A:Residues: 1-427 <HVA>
A:Cross-references: FIDB:XA1800; MIMD:3167779; FIDB:AAK2990.1; FIDB:G142779
A:Accession: S78114
A:Molecule type: protein
A:Residues: 10,85,96,110,140,165,174,210,221,232,249,277,285,293,309,337,390,430-425-
R:Kim, D.H.; Kaneko, C.; Mizokami, Y.
Biochim. Biophys. Acta 1122, 203-211, 1992
A:Title: Purification and characterization of major glycoproteins, PAS-6 and PAS-7, I
A:Reference number: S24181; MIMD:92353107
A:Accession: S24181
A:Molecule type: protein
A:Residues: 383-394 <KIM>
C:Superfamily: milk fat globule protein, discoidin I amino-terminal homology; EGF hom
C:Keywords: blocked amino end; disulfide bond; glycoprotein; milk
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-427/Product: PAS-6/7 protein #status experimental <MVS>
F:24-59/Domain: EGF homology <EG>
F:66-105/Domain: EGF homology <EG>
F:108-255/Domain: discoidin I amino-terminal homology <DN>

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #next_change 04-Mar-2000
 C:Accession: T27674
 R:Mortimore, B.
 Submitted to the EMBL Data Library, August 1994
 A:Reference number: Z20403
 A:Accession: T27674
 A:Status: preliminary; translated from GB/EMBL/DDHJ
 A:Molecule type: DNA
 A:Residues: 1744 <N1>
 A:Cross-references: EMBL:Z20403; FIDN:CAAR4676.1; GSPDR:G000021; CESP:ZK1058.1
 A:Experimental source: cDNA ZK1058
 C:Genetics:
 A:Gene: CESP:ZK1058.1
 A:Map position: 3
 A:Indices: 1/2 35/73, 439/73, 500/73, 697/73
 C:Superfamily: Human methylmalonyl-CoA mutase; cobalamin-binding homology

Query Match 73.9% Score 34; DB 2; Length 744;
 Best Local Similarity 75.0%; Pred. No. 65;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Q7 2 LNFETVEA 9
 |||||
 Db 64 LNFETVEA 64

RESULT 14
 T50155
 hypothetical protein SHAC222.145 [imported] fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C>Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #next_change 21-Jul-2000
 C:Accession: T50155
 R:Hamlin, N.; Churcher, C.M.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
 Submitted to the EMBL Data Library, November 1999
 A:Reference number: Z25043
 A:Accession: T50155
 A:Status: preliminary; translated from GB/EMBL/DDHJ
 A:Molecule type: DNA
 A:Residues: 1762 <N1>
 A:Cross-references: EMBL:Z25043; FIDN:AA0766.1; GSPDR:G00000; SRR:SPN1722.1
 A:Experimental source: strain 972h() cosmid c222
 C:Genetics:
 A:Gene: SPDB:SPAC222.14c
 A:Map position: 1
 A:Indices: 53/71, 78/71, 106/73
 C:Superfamily: Saccharomyces cerevisiae probable membrane protein YOR165w

Query Match 73.9% Score 34; DB 2; Length 762;
 Best Local Similarity 66.7%; Pred. No. 67;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Q7 1 LNFETVEA 9
 |||||
 Db 328 LNFETVEA 336

RESULT 15
 E81282
 probable periplasmic protein Cj1376 [imported] Campylobacter jejuni (strain NC9C 11)
 C:Species: Campylobacter jejuni
 C>Date: 01-Mar-2000 #sequence_revision 01-Mar-2000 #next_change 01-Mar-2000
 C:Accession: E81282
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kettle, J.M.; Churcher, C.; Basham, D.; Chit
 Nature 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
 A:Reference number: AB1250; MUID:20159912
 A:Accession: E81282
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 269 74P

Query Match 73.9% Score 34; DB 2; Length 409;
 Best Local Similarity 75.0%; Pred. No. 33;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Q7 1 LNFETVEA 8
 |||||
 Db 215 LNFETVEA 223

RESULT 13
 T27674
 hypothetical protein ZK1058.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans

Q7 1 LNFETPVEA 9
 |||||
 Db 66 LNFETPVEA 74

RESULT 11
 D69267
 ubiquinol:coenzyme A:acyl-CoA synthetase (UBS) homolog - Archaeoglobus ful
 C:Species: Archaeoglobus fulgidus
 C>Date: 05-Jun-1997 #sequence_revision 05-Jun-1997 #next_change 21-Jul-2000
 C:Accession: D69267
 R:Klenk, H.F.; Clayton, K.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 Glueck, A.; Zhou, J.; Gocke, B.; Goggin, T.; White, T.F.; Miller, L.L.
 Nature 390, 344-370, 1997
 A:Authors: Uitterlinden, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Wiese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate reducing archaeo
 A:Reference number: A69250; MUID:98049343
 A:Accession: D69267
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-261 <N1>
 A:Cross-references: EMBL:Z69267; FIDN:AA0782.1; GSPDR:G00001; FID:925043
 C:Superfamily: bioc homology
 F:61-159/domain: bioc homology <BIOC>

Query Match 73.9% Score 24; DB 2; Length 261;
 Best Local Similarity 75.0%; Pred. No. 20;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Q7 1 LNFETVEA 9
 |||||
 Db 35 LNFETVEA 42

RESULT 12
 T11743
 PP47 protein - pig (import)
 C:Species: Sus scrofa domestica (domestic pig)
 C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #next_change 21-Jul-2000
 C:Accession: T11743
 R:Busch, M.; Vogel, T.; Calvete, J.J.; Thiele, H.H.; Schmidtke, J.; Matsuda, T.; Toepfe
 Biol. Reprod. 58, 1057-1064, 1998
 A:Title: Molecular cloning and characterization of P47, a novel boar sperm-associated co
 A:Reference number: Z17325; MUID:98206817
 A:Accession: T11743
 A:Status: preliminary; translated from GB/EMBL/DDHJ
 A:Molecule type: mRNA
 A:Residues: 1-409 <N1>
 A:Cross-references: EMBL:Y11683; FIDN:CAAT2373.1; FID:9252928
 A:Experimental source: testis
 C:Function:
 A:Description: may be involved in membrane remodeling and/or function as a zona pellucid
 F:6-10/Domain: EGF homology <EGF>

Query Match 73.9% Score 34; DB 2; Length 409;
 Best Local Similarity 75.0%; Pred. No. 33;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Q7 1 LNFETVEA 8
 |||||
 Db 215 LNFETVEA 223

RESULT 13
 T27674
 hypothetical protein ZK1058.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans

A:Cross-References: GR:AL149078; GH:AL111168; NID:96968723; FIUN:CA873893.1; PID:9696884
 A:Experimental source: serotype 02, strain NCTC 11168
 C:Genetics:
 A:Gene: Cj1376

Query Match 71.7%; Score 33; DB 2; Length 269;
 Best Local Similarity 75.0%; Pred. No. 34;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 NLEFETPE 8
 DB 80 NLENTPVK 87

Search completed: September 5, 2002, 15:28:52
 Job time: 457 sec



Genome version 4.5
Copyright (c) 1993 - 2000 Genome Inc.

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:31:41 ; Search time 33.99 Seconds
(without alignments)
10.252 Million cell updates/sec

Title: US-09-744-804-38
Perfect score: 46
Sequence: 1 NLFETPVEA 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 24

Listing first 45 summaries

Database: SwissProt_40 *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	%	ID	Description
1	46	100.0	387	1 MFGM_HUMAN	Q08431 homo sapien
2	38	82.6	2261	1 ARO1_HUMAN	Q65477 homo sapien
3	38	82.6	2261	1 ARO1_MOUSE	P11233 mus musculu
4	36	78.3	427	1 MFGM_BOVIN	Q96114 bos taurus
5	34	73.9	409	1 MFGM_PIG	F79385 sus scrofa
6	34	73.9	744	1 MFGM_CAMEL	Q23381 camorhabdi
7	33	71.7	293	1 BLAC_FRO-7A	P14171 rhodobacter
8	33	71.7	464	1 YSV5_CAMEL	Q10011 camorhabdi
9	33	71.7	535	1 SPRC_SYNY3	P74745 synchocyst
10	33	71.7	612	1 NU5M_GADMO	P55782 gadus morhu
11	32	69.6	291	1 Y12K_SSV1	P20197 solfobacta
12	32	69.6	312	1 GUB2_HORVU	P12257 hordeum vul
13	32	69.6	486	1 Y6PM_YEPHE	P17778 yeastia pe
14	32	69.6	744	1 NUB5_YEAST	P46673 saccharomyc
15	32	69.6	1363	1 VGR3_MOUSE	P35317 mus muscula
16	31	67.4	62	1 RSC1_ANASP	P49224 anabacta sp
17	31	67.4	217	1 GSEL_HUMAN	Q46877 homo sapien
18	31	67.4	217	1 GSEL_PAT	P29576 rattus norv
19	31	67.4	255	1 Y688_HELPJ	Q92833 helicobacte
20	31	67.4	255	1 Y688_HELPJ	Q92833 helicobacte
21	31	67.4	260	1 MFPY_PACSB	Q57332 bacillus su
22	31	67.4	334	1 PEUR_RACSP	P40410 bacillus so
23	31	67.4	475	1 YDFG_SCHFC	Q16478 schizosach
24	31	67.4	422	1 GATR_ARCFU	Q28164 archaeflaba
25	31	67.4	581	1 LEPA_MYCPN	P75498 mycoplasma
26	31	67.4	548	1 LEPA_MYCPN	P47484 mycoplasma
27	31	67.4	665	1 SINT_SCHPO	Q97799 schizosach
28	31	67.4	712	1 MUTE_RHIME	Q86528 rhizobium m
29	31	67.4	715	1 MUTE_PORC1	Q59677 porcyporom
30	31	67.4	825	1 ILAR_HUMAN	P24394 homo sapien
31	31	67.4	1385	1 FAT1_HUMAN	Q14745 schizosach
32	30	65.2	176	1 Y197_ARCFU	Q29800 archaeflaba
33	30	65.2	190	1 NUDB_XYIFA	Q99949 xyfella fas

RESULT	1	34	30	65.2	231	1	ALSP_ECOLI
MFGM_HUMAN							
ID	MFGM_HUMAN						
AC	Q08431						
DT	01-OCT-1996						
DT	01-NOV-1997						
DT	16-OCT-2001						
DE	Lactadherin precursor (Milk fat globule EGF factor 8) (MFG-E8) (HMEG)						
DE	(Breast epithelial antigen EAM6) (MEGM) (Contains: Medin)						
GN	MEGE8						
OS	Homo sapiens (Human)						
OC	Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;						
OC	Mammalia, Euthalia, Primates, Catarrhini, Hominidae; Homo.						
OX	NCBI_TaxID=9606;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
AC	ISSUE-Breast, and Breast carcinoma;						
RX	MEDLINE=96213908; PubMed=8639264;						
PA	Couto J.P., Taylor M.P., Godwin S.G., Ceriani P.L., Peterson J.A.;						
RT	"Cloning and sequence analysis of human breast epithelial antigen						
RT	RAM6 reveals an EGF cell adhesion sequence presented on an epidermal						
RT	growth factor-like domain."						
RT	DNA Cell Biol. 15:281-286(1996).						
FN	[2]						
FP	SEQUENCE OF 176-187 FROM N.A.						
FC	TISSUE Mammary gland;						
FX	MPIIMP 9137155; PubMed 1995932;						
FA	Larocca R., Peterson J.A., Upton F., Kuniyoshi J., Bistrain A.M.,						
FA	Ceriani P.L.;						
FI	"A Mr 42,000 human milk fat globule protein that is highly expressed						
FI	in human breast tumors contains factor VIII-like domains."						
FN	[3]						
FP	PARTIAL SEQUENCE, AND CHARACTERIZATION.						
FC	TISSUE Milk;						
FX	MEDLINE 96134924; PubMed 9595276;						
FA	Giuffrida M.C., Cavalletto M., Giunta C., Conti A.,						
FA	Gedevicius Rimantas J.;						
FT	"Isolation and characterization of full and truncated forms of human						
FT	breast carcinoma protein RAM6 from human milk fat globule membranes."						
FI	[4]						
FN	Protein Chem. 17:143-148(1998).						
FP	SEQUENCE OF 665-17, AND IDENTIFICATION OF MEDIN.						
FX	MEDLINE 99342076; PubMed 10411933;						
FA	Hargreaves R., Nussland T., Sletten K., Westmark G.T., Muschiano G.,						
FA	Tjoharj L.G., Nordsted C., Engstrom B., Westmark P.;						
FI	"Medin, an internal fragment of active smooth muscle cell-produced						
FI	Lactadherin forms the most common human amyloid."						
FN	[5]						
FP	CHARACTERIZATION.						
FX	MEDLINE=97455885; PubMed=9280929;						
FA	Taylor M.P., Couto J.P., Scallan C.D., Ceriani P.L., Peterson J.A.;						
FA	"Lactadherin (formerly RAM6), a membrane-associated glycoprotein						
FI	expressed in human milk and breast carcinomas, promotes A11-Gly-Asp						
FI	(PGN)-dependent cell adhesion."						

ALIGNMENTS

ID	MFGM_HUMAN	STANDARD;	PRT;	387 AA.
AC	Q08431			

DT 01-OCT-1996 (Pel. 34, Created)

DT 01-NOV-1997 (Pel. 35, Last sequence update)

DT 16-OCT-2001 (Pel. 40, Last annotation update)

DE Lactadherin precursor (Milk fat globule EGF factor 8) (MFG-E8) (HMEG)

DE (Breast epithelial antigen EAM6) (MEGM) (Contains: Medin)

GN MEGE8

OS Homo sapiens (Human)

OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi;

OC Mammalia, Euthalia, Primates, Catarrhini, Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

AC ISSUE-Breast, and Breast carcinoma;

RX MEDLINE=96213908; PubMed=8639264;

PA Couto J.P., Taylor M.P., Godwin S.G., Ceriani P.L., Peterson J.A.;

RT "Cloning and sequence analysis of human breast epithelial antigen

RAM6 reveals an EGF cell adhesion sequence presented on an epidermal

growth factor-like domain."

RT DNA Cell Biol. 15:281-286(1996).

FN [2]

FP SEQUENCE OF 176-187 FROM N.A.

FC TISSUE Mammary gland;

FX MPIIMP 9137155; PubMed 1995932;

FA Larocca R., Peterson J.A., Upton F., Kuniyoshi J., Bistrain A.M.,

Ceriani P.L.;

FI "A Mr 42,000 human milk fat globule protein that is highly expressed

in human breast tumors contains factor VIII-like domains."

FN [3]

FP PARTIAL SEQUENCE, AND CHARACTERIZATION.

FC TISSUE Milk;

FX MEDLINE 96134924; PubMed 9595276;

FA Giuffrida M.C., Cavalletto M., Giunta C., Conti A.,

Gedevicius Rimantas J.;

FT "Isolation and characterization of full and truncated forms of human

breast carcinoma protein RAM6 from human milk fat globule membranes."

FI [4]

FN SEQUENCE OF 665-17, AND IDENTIFICATION OF MEDIN.

FX MEDLINE 99342076; PubMed 10411933;

FA Hargreaves R., Nussland T., Sletten K., Westmark G.T., Muschiano G.,

Tjoharj L.G., Nordsted C., Engstrom B., Westmark P.;

FI "Medin, an internal fragment of active smooth muscle cell-produced

Lactadherin forms the most common human amyloid."

FN [5]

FP CHARACTERIZATION.

FX MEDLINE=97455885; PubMed=9280929;

FA Taylor M.P., Couto J.P., Scallan C.D., Ceriani P.L., Peterson J.A.;

"Lactadherin (formerly RAM6), a membrane-associated glycoprotein

expressed in human milk and breast carcinomas, promotes A11-Gly-Asp

(PGN)-dependent cell adhesion."

p32719	escherichia
p12745	halobacter
p73335	synechocyst
Q04543	simian vari
Q97828	helicobacte
Q25801	helicobacte
Q12946	platichtys
Q02558	aspergillus
P08414	mus musculu
P12337	homo sapien
P12334	rattus norv
Q09417	schizosach

RC STRAIN=DNA72; TISSUE=MASTOEPITHELIAL;
 RX MEDLINE=94375004; PubMed=8086782;
 PA Luciani M F, Fendler F, Savazzi S, Mollat M, et al, Chaminet P;
 RT "Cloning of two novel ABC transporters mapping on human chromosome
 9";
 RL Genomics 21:150-154(1994)
 RN [2]
 RP SEQUENCE FROM N A
 RC STRAIN=C57BL/6J;
 RA Qiu Y, Cavalier L, Chiu S, Rubin F, Cheng J, et al;
 RT "Human and mouse ABCA1 comparative sequencing and transgenesis studies
 identify potential regulatory sequences";
 RL Submitted (Jul-2000) to the EMBL/GenBank/DBP1 databases
 CC [-] FUNCTION: CAMP-DEPENDENT AND SULFONYLUREA-SENSITIVE ANION
 CC TRANSPORTER. KEY GATEKEEPER INFLUENCING INTRACELLULAR CHLORIDE-LEVEL.
 CC TRANSPORT (BY SIMILARITY).
 CC [-] TISSUE SPECIFICITY: WIDELY EXPRESSED IN ADULT TISSUES. HIGHEST
 CC LEVELS ARE FOUND IN PREGNANT UTERUS AND UTERUS
 CC [-] DOMAIN: MULTIFUNCTIONAL POLYPEPTIDE WITH TWO HOMOLOGOUS HALVES,
 CC EACH CONTAINING AN HYDROPHOBIC MEMBRANE-ANCHORING DOMAIN AND AN
 CC ATP BINDING CASSETTE (ABC) DOMAIN.
 CC [-] SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. ABCA SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, except by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements/>
 CC or send an email to license@sib.sib.ch).
 CC
 DR EMBL: X75926; CAA53530.1; ALT_INT.
 DR EMBL: AF287263; AAC39073.1; ALT_INT.
 DR MGD: MGI:99607; Abca1.
 DR InterPro: IPR003493; AAA
 DR InterPro: IPR003439; ABC transporter.
 DR InterPro: IPR001687; ALT_INT_A
 DR InterPro: IPR003838; DUF214.
 DR InterPro: IPR000897; SRP54.
 DR Pfam: PF00005; ABC_tran; 2.
 DR Pfam: PF02687; DUF214; 1.
 DR Pfam: PF00448; SRP54; 1.
 DR SMART: SMO0382; AAA; 1.
 DR PROSITE: PS00211; ABC_TRANSPORTER; 1.
 KW ATP binding, Glycoprotein, Transmembrane; Transport.
 FT TRANSMEM 26 42 POTENTIAL;
 FT TRANSMEM 640 656 POTENTIAL;
 FT TRANSMEM 690 706 POTENTIAL;
 FT TRANSMEM 717 733 POTENTIAL;
 FT TRANSMEM 749 765 POTENTIAL;
 FT TRANSMEM 771 787 POTENTIAL;
 FT TRANSMEM 1041 1057 POTENTIAL;
 FT TRANSMEM 1351 1367 POTENTIAL;
 FT TRANSMEM 1661 1677 POTENTIAL;
 FT TRANSMEM 1708 1724 POTENTIAL;
 FT TRANSMEM 1737 1753 POTENTIAL;
 FT TRANSMEM 1775 1791 POTENTIAL;
 FT TRANSMEM 1854 1870 POTENTIAL;
 FT NP_BIND 933 940 ATP (POTENTIAL).
 FT NP_BIND 1946 1953 ATP (POTENTIAL).
 FT CARBOHYD 14 14 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 98 98 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 151 151 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 161 161 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 166 166 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 244 244 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 292 292 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 337 337 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 349 349 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 400 400 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 478 478 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 489 489 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 521 521 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 829 829 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1144 1144 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1294 1294 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1453 1453 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1499 1499 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1504 1504 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1637 1637 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 2044 2044 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 2248 2248 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1567 1568 MISSING (IN REF. 2)
 FT CONFLICT 2024 2024 MISSING (IN REF. 2)
 SQ SEQUENCE 4261 AA; 254311 MW: 546542; ED1005F9 GRC64;
 Query Match: 82.6%, Score 59, Id 11, Length 4261;
 Best Local Similarity 87.5%; Pred. No. 10;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NLFFPVR 8
 DQ 808 NLFFPVR 815
 RESULT 4
 MEGM_BOVIN
 ID MEGM_BOVIN STANDARD; PPT: 427 AA.
 AC Q95114; Q27959; P79344;
 DT 01-NOV-1997 (Ref 35, Created)
 DI 01-NOV-1997 (Ref 35, Last sequence update)
 DE Lactadherin precursor (Milk fat globule-BF factor 8) (MGF-BF)
 DE (MGF57/53) (PAS-6/PAS-7 glycoprotein) (MGFM) (Sperm surface protein
 DE SP47) (BP47) (Components 15/16).
 GN MEGE8
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Puminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 CX NCBI_TaxID=9913;
 IL [1]
 RF SEQUENCE FROM N A. PARTIAL SEQUENCE, AND CARBOHYDRATE-LINKAGE SITES.
 RC STRAIN=HOLSTEIN; TISSUE=Mammary gland;
 RX MEDLINE=97008954; PubMed=8856064;
 PA Hvarregaard J, Andersen M H, Berglund J, Easmussen J T,
 RA Petersen T E;
 RT "Characterization of glycoprotein PAS-6/7 from membranes of bovine
 RT milk fat globules";
 RL Eur. J. Biochem. 249:628-636(1996).
 RN [2]
 RP SEQUENCE OF 18-427 FROM N A.
 RC TISSUE=Mammary gland;
 RX MEDLINE=96125736; PubMed=8541316;
 PA Aoki N, Kishi M, Taniguchi Y, Adachi T, Nakamura K,
 RA Matsuda T;
 RT "Molecular cloning of glycoprotein oligosaccharide recognized by
 RT monoclonal antibodies raised against bovine milk fat globule
 RT membrane";
 RL Biochim. Biophys. Acta 1245:385-391(1995).
 RN [3]
 RP SEQUENCE OF 18-427 FROM N A.
 RC TISSUE=Testis;
 RX Fossili M A;
 PA submitted (N-V-1997) to the EMBL/GenBank/DBP1 databases.
 RN [4]
 RP SEQUENCE OF 140-146; 174-187; 194-245 AND 422-427
 RC TISSUE=Milk;
 RX MEDLINE=93250576; PubMed=8485470;
 PA Marier I H, Raghart P, Lane W S;
 RT "The major fat-globule membrane proteins, bovine components 15/16 and
 RT guinea-pig 55, are homologous to MGF-BF, a murine glycoprotein
 RT containing epidermal growth factor-like and factor V/VIII-like
 RT sequences";
 RL Biochem. Mol. Biol. Int. 24:545-554(1993).


```

RESULT 6
MUTA_CAEEL
ID MUTA_CAEEL STANDARD; PRT; 744 AA.
AC Q238L;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable methylmalonyl-CoA mutase, mitochondrial precursor
DE (EC 5.4.99.2) (MCM).
GN ZK1058.1
OS Caenorhabditis elegans.
OC Eukarya, Metazoa, Nematoda, Chromadorea, Rhabditida, Rhabditoidea.
OC Rhabditidae, Peloderinae, Caenorhabditis
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Mortimore B.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases
CC -!- FUNCTION: INVOLVED, IN MAN, IN THE DEGRADATION OF SEVERAL AMINO
ACIDS, GLO-CHAIN FATTY ACIDS AND CHOLESTEROL VIA PROXYMETHYL-COA TO
CC THE TRICARBOXYLIC ACID CYCLE. MCM HAS DIFFERENT FUNCTIONS IN
CC OTHER SPECIES (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: (F) 2-methyl-3-oxopropionyl-CoA succinyl-
CoA.
CC -!- COFACTOR: ADENOSYLCOBALAMIN (BY SIMILARITY).
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix (by similarity).
CC -!- SIMILARITY: BELONGS TO THE METHYLMALONYL-COA MUTASE FAMILY.
CC
CC This SWISS-Prot entry is copyright © It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announc/
CC or send an email to license@isb-sib.ch).
CC
FMRI: 735604; CAAR4676.1; -
DR HSSP; P11653; LPEQ
DR WormRep; ZK1058.1; CE01101.
DR InterPro; IPR003312; R12-binding.
DR Pfam; PF02310; R12 binding; 1
DR Pfam; PF01642; MM_CoA_mutase; 1
DR PROSITE; PS00544; METHMALONYL_COA_MUTASE; 1.
KW Mitochondrion; Transit peptide; Isomerase; Vitamin B12; Cobalt.
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN ? 744 PROBABLE METHYLMALONYL-COA MUTASE.
FT DOMAIN 616 694 COBALAMIN-BINDING (POTENTIAL).
FT METAL 618 618 COBALT (POTENTIAL).
SQ SEQUENCE 744 AA; 81669 MW; 4871e16a93c0f206 CRC64;

Query Match 73.9%; Score 34; DB 1; Length 744;
Best Local Similarity 75.9%; Pred No 22;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LFETPEVA 9
1111111
DB 641 LFQTPLEA 648

RESULT 7
BLAC_RHOCA
ID BLAC_RHOCA STANDARD; PRT; 293 AA.
AC P14171;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Beta-lactamase precursor (EC 3.5.2.6) (Penicillinase).
OS Rhodobacter capsulatus (Rhodospirillum rubrum)
OC Bacteria, Proteobacteria, alpha sub-division, Rhodobacter group;

```

```

Rhodobacter.
OC NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RA Campbell J. I. A.; Seebull S.; Gillespie T.; Ankley P. P.;
DE "The phototrophic bacterium Rhodospirillum rubrum encodes
DE an indigenous class A beta-lactamase.";
DE Biochem J 260:803-812(1990).
CC -!- FUNCTION: HYDROLYSES BETA-LACTAMS ANTIBIOTICS. RATES OF
CC HYPERLYSIS RELATIVE TO BENZYL-PENICILLIN 1000 AMPCILLIN = 97,
CC CEFEPIME = 25, CLOXACILIN = 9, CEPHALORIDINE = 4.
CC -!- CATALYTIC ACTIVITY: A beta lactam + H(2)O = a substituted beta-
amino acid.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- SIMILARITY: BELONGS TO THE CLASS-A BETA-LACTAMASE FAMILY.
CC
CC This SWISS-Prot entry is copyright © It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announc/
CC or send an email to license@isb-sib.ch).
CC
FMRI: X15791; CAAR3795.1; -
DR PIR; S04649; S04649.
DR HSSP; P00810; ITEM.
DR InterPro; IPR001466; Beta_lactam.
DR InterPro; IPR000871; Beta_lactam_A.
DR Pfam; PF00144; beta-lactamase; 1.
DR PRINTS; PR00119; BLACTAMASE.
DR PROSITE; PS00146; BETA-LACTAMASE_A; 1.
KW Hydrolyase, Antibiotic resistance, Periplasmic, Signal.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 293 BETA-LACTAMASE.
FT ACT_SITE 74 74 BY SIMILARITY.
FT BINDING 238 240 SUBSTRATE (BY SIMILARITY).
SQ SEQUENCE 293 AA; 41005 MW; 40925906a58017f CRC64;

Query Match 71.79%; Score 33; DB 1; Length 293;
Best Local Similarity 77.8%; Pred No 14;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLETPVEA 9
1111111
DB 25 LAETPEVA 33

RESULT 8
YSV5_CAEEL
ID YSV5_CAEEL STANDARD; PRT; 464 AA.
AC Q10011;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hydrophobic 61.3 kDa protein T1935 in Chromosome III precursor.
GN T19C3.5.
OS Caenorhabditis elegans.
OC Eukarya, Metazoa, Nematoda, Chromadorea, Rhabditida, Rhabditoidea.
OC Rhabditidae, Peloderinae, Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Fulton L.;
RL Submitted (JUN 1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: SAME, TO THE PELTIFIP/TRE/PLTP FAMILY.
CC
CC This SWISS-Prot entry is copyright © It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement>
 CC or send an email to license@sib-sib.ch).

CC EMBL: U29412; AAC46506.1; -
 DR WormPep: T19C3.5; CMO2056
 DR InterPro: IPR001124; IPR_RPL_CETP.
 DR Pfam: PF01273; IPR_RPL_CETP_1.
 DR Pfam: PF02886; IPR_RPL_CETP_C_1.
 DR SMART: SM00328; BP11; 1.
 DR SMART: SM00329; BP12; 1.
 DR ProSite: PS00400; LRP_RPL_CETP; FALSE_NEG.
 KW Hypothetical protein, Signal.
 FT SIGNAL 1 17 POTENTIAL
 FT CHAIN 18 464 HYPOTHETICAL PROTEIN T19C3.5.
 SQ SEQUENCE 464 AA: 51939 MW: 30290CD72F65E67F CRC64;

Query Match 71.7%; Score 33; DB 1; Length 464;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 1 NLFETP 6
 |||||
 Db 134 NLFETP 139

RESULT 9
 SPKC_SYNY3 STANDARD; PRT: 535 AA
 AC P74745;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable serine/threonine-protein kinase C (EC 2.7.1.1).
 GN SPKC OR SLR0599.
 OS Synecocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chlorococcales; Synecocystis.
 CC NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kamei A., Ikeuchi M.;
 FT "A novel gene, *skcC*, encodes active Ser/Thr protein kinase in the
 FT motile cyanobacterium *Synechocystis* sp. PCC 6803";
 RL Submitted (JUL-2000) to the EMBL/GenBank/Trna databases
 RN [2]

CC SEQUENCE FROM N.A.
 EX MEDLINE 97361201, PubMed:8905231;
 PA Kaneke T., Sato S., Koyani H., Tanaka A., Asanizu E., Nakamura Y.,
 PA Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,
 PA Hosouchi T., Matsuno A., Moraki A., Nakazaki N., Niyuu K., Okumura S.,
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 CC "Sequence analysis of the genome of the unicellular cyanobacterium
 RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions";
 RL DNA Res 3:106-136(1996)
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: AB046599; BAB17035.1; -
 DR EMBL: D40417; RAAL0865.1;
 DR InterPro: IPR000719; Euk_thr_kinase.
 DR InterPro: IPR002203; Ser_thr_kinase.

IP Pfam: PF00069; pkinase.1
 LR ProSite: PS00611; PROTEIN_KINASE_DOM; 1.
 LR ProSite: PS00107; PROTEIN_KINASE_ATP; 1.
 LR ProSite: PS00108; PROTEIN_KINASE_ST; 1.
 KW Transferase, Serine/threonine protein kinase, ATP binding;
 KW Complete proteome.
 FT DOMAIN 12 277 PROTEIN KINASE.
 FT NP_BIND 18 26 ATP (BY SIMILARITY).
 FT BINDING 43 43 ATP (BY SIMILARITY).
 FT ACCL_SITE 112 142 BY SIMILARITY.
 SQ SEQUENCE 535 AA: 58141 MW: 53305F30E325D8 CRC64;

Query Match 71.7%; Score 33; DB 1; Length 535;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 1 NLFETP 6
 |||||
 Db 396 NLFETP 401

RESULT 10
 NU5M_GADMO STANDARD; PRT: 612 AA.
 ID NU5M_GADMO
 AC P5782;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE NADH-ubiquinone oxidoreductase chain 5 (EC 1.6.5.3).
 GN MTND5 OR ND5.
 OS *Gadus morhua* (Atlantic cod).
 OC Mitochondrion.
 CC Eukaryota; Metazoa; Chordata; Clariata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 CC Acipenseriformes; Paracanthopterygii; Gadiformes; Gadidae;
 CC *Gadus*.
 CC NCBI_TaxID=8049;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NOHFWGIAN COASTAL 1;
 FX MEDLINE=9441975; PubMed=8617926;
 RA Johansen S., Bakke I.;
 FT "The complete mitochondrial DNA sequence of Atlantic cod (*Gadus*
 FT *medhua*), relevance to taxonomic studies among codfishes";
 RL Mol. Mar. Biol. Biotechnol. 5:203-214(1996).
 RN -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.

CC This SWISS-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@sib-sib.ch).

CC EMBL: X99772; CAA68115.1; -
 DR InterPro: IPR003916; NADH_oxidoreductse5.
 DR InterPro: IPR001750; Oxidored_q1.
 DR InterPro: IPR001516; Oxidored_q1_N.
 DR Pfam: PF00361; Oxidored_q1_1.
 DE Pfam: PF00572; Oxidored_q1_N; 1.
 DE PF01434; NADH_OXIDASE5.
 KW Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
 SQ SEQUENCE 612 AA: 66449 MW: 14639P2P51F1A159 CRC64;

Query Match 71.7%; Score 33; DB 1; Length 612;
 Best Local Similarity 85.7%; Pred. No. 30;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0.

QY 1 NLFETPV 7
 |||||

OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
PC STRAIN-KIM5;
FA MEDLINE=8945690; PubMed=2670848;
RX Leung K.Y., Straley S.C.;
RT "The ygm gene of *Yersinia pestis* encodes a released protein having
PT homology with the human platelet surface protein GpIIb/alpha";
EL J. Bacteriol. 171:4523-4532(1989).
RN [2]
RP SEQUENCE FROM N.A.
PC STRAIN-KIM5;
RX MEDLINE=94427122; PubMed=2745557;
PA Perry P.P., Straley S.C., Fetherston J.P., Rose P.J., Gregor J.,
Blattner F.R.;
RT "RNA sequencing and analysis of the low Cu2+-response plasmid pCD1 of
PT *Yersinia pestis* KIM5";
EL Infect Immun 66:4611-4623(1998).
RN [3]
RP SEQUENCE FROM N.A.
PC STRAIN-KIM5;
RX MEDLINE=94424474; PubMed=2748454;
PA Hu P., Elliott J., McCready P., Skowronski E., Barnes J.,
Fukuyoshi A., Rubaker F.F., Garcia E.;
RT "Structural organization of virulence-associated plasmids of *Yersinia*
RT *pestis*";
EL J. Bacteriol. 180:5192-5202(1998).
RN [4]
RP SEQUENCE FROM N.A.
PC STRAIN-CO-92 / Biovar Orientalis;
RX Medline=21470413; PubMed=11586360;
FA Zuckhill J., Wilson R.W., Thompson N.F., Tibbitt F.W., Bisher M.L.B.,
RA Prentice M.B., Sebahia M., James K.D., Chutcher C., Bungail K.L.,
BA Baker S., Pesham D., Bentley S.D., Brooks K., Carden Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies F.M., Davis F., Bangay R.,
RA Feltwell T., Hamlin N., McIlroy S., Jagels K., Karlyshev A.V.,
RA Leather T., Moule S., Gyston P.C.F., Quail M.,utherford K.,
PA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genomic sequence of *Yersinia pestis*, the causative agent of plague";
EL Nature 413:523-527(2001).
PL NCBI_TaxID=632;
CC -!- SUBCELLULAR LOCATION: OUTER MEMBRANE AND SECRETED.
CC -!- SIMILARITY: CONTAINS 12 LEUINF-FICH REPEATS (LFF)
CC -!- SIMILARITY: TO S.FLEXNERI IPAH 7.8
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/faq/faq.html>
CC or send an email to license@sib-sib.ch)
CC
EMBL: M25810; AAA27670.1;
DR EMBL: AF074612; AA265806.1;
DR EMBL: AF053946; AA062580.1;
DR EMBL: AL117189; CAB54903.1;
DR PIR: A33950; A33950.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR005892; LRR_out
DR Pfam: PF00560; LRR; 6.
DR SMART: SM00370; LRR; 12
DR CATH: 3.40.130.10; LRR; 12
KW Cysteine repeats, Repeat, Domain, Fish repeat, Antigen, Plasmid,
KW Complete proteome.
FT FEAT 59 56 LRR 1
FT REPEAT 91 110 LRR 2
FT REPEAT 111 129 LRR 3
FT REPEAT 130 152 LRR 4
FT REPEAT 153 177 LRR 5
FT REPEAT 179 194 LRR 6
FT REPEAT 195 217 LRR 7
FT REPEAT 236 256 LRR 8
FT PEPAT 257 281 LPP 9
FT REPEAT 296 316 LPP 10

FT REPEAT 317 341 LRR 11
FT REPEAT 356 381 LRR 12
FT CONFLICT 142 183 MISSING (IN REF. 1).
SQ SEQUENCE 409 AA, 4529 MW, 122400.671 AC0951 CP064;

Query Match 59 56 Score 32; DB 1; Length 409;
Best local similarity 77 84; Pos: 10; 32;
Matches: 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLFETPVEA 9
||| ||||
DL 22 NUTEMPVEA 30

RESULT 14
ID N085_YEAST STANDARD; PRT; 744 AA.
AC P46673;
DE 01-NOV-1995 (Rel. 42, Created)
DI 01-NOV-1995 (Rel. 32, Last sequence update)
DE 30 MAY 2005 (Rel. 59, Last annotation update)
DE Nucleoporin NUP85 (Nuclear pore protein NUP85)
GN NUP85 OF RA15 OF YK042W OF J1624.
OS *Saccharomyces cerevisiae* (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetia; Saccharomycetes;
OC Saccharomycetidae; Saccharomycetaceae; Saccharomycetaceae.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
PC STRAIN-FY23;
FA MEDLINE=9443362; PubMed=887938;
RA Goldstein A.L., Day C.A., Heath C.V., Goldstein N.;
RT "Phenotypic nuclear defects associated with a conditional allele of
PL the gene nucleoporin Nup85p";
EL Mol Biol Cell 7:917-934(1996).
RN [2]
RP SEQUENCE FROM H.A.; AND SEQUENCE OF 31 39; 123 138 AND 708-718.
FA MEDLINE=961256; PubMed=8565072;
RX Strickson S., Wimmer C., Fieger M., Doye V., Tekotte H., Weise C.,
FA Emig S., Sedgwick A., Hurt E.C.;
RT "A novel complex of nucleoporins, which includes Nup85 and a Nup85p
PT homolog, is essential for normal nuclear pores";
EL Cell 84:265-275(1996).
RN [3]
RP SEQUENCE FROM N.A.
PC STRAIN-S788C;
FA MEDLINE=9539755; PubMed=766047;
RX Huang M.-F., Chant J.-C., Gallibert F.;
RT "Analysis of a 42.5 kb cDNA sequence of chromosome X reveals three
PL cDNA genes and 14 new open reading frames including a gene most
RT probably belonging to the family of ubiquitin-protein ligases";
EL Yeast 11:775-781(1995).
CC -!- FUNCTION: PROPOSED FOR EFFICIENT mRNA EXPORT FROM THE NUCLEUS TO
CC THE CYTOSOL AND FOR CORRECT NUCLEAR PORE BIOGENESIS.
CC -!- SUBUNIT: INTERACTS WITH NUP84, NUP120, NUP133 AND NUP155.
CC -!- SUBCELLULAR LOCATION: Nuclear pore complex.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/faq/faq.html>
CC or send an email to license@sib-sib.ch)
CC
EMBL: M36469; AA048443.1;
DR EMBL: X90995; CA062481.1;
DR EMBL: Z49542; CA089569.1;
DR EMBL: I36344; AA089744.1;
DR SGD: S0003803; NUP85.
KW Nuclear protein, Transport.
FT CONFLICT 181 191 S 7 T (IN REF. 1).

FT CONFLICT 471 471 F -- L (IN REF. 1).
 SQ SEQUENCE 744 AA 84897 MW 40944AD3G857A023 CR664.

 Query Match 69.6% Score 32; DB 1; Length 744;
 Best Local Similarity 66.7% Pred. No. 59;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 QY 1 NLEPTVEA 9
 |||||
 Db 708 NLEPTVEA 716

 RESULT 15
 VGR3_MOUSE STANDARD; PRI: 1363 AA.
 AC P45917;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Vascular endothelial growth factor receptor 3 precursor (EC 2.7.1.112)
 DE (VEGFR-3) (tyrosine-protein kinase receptor FLT4)
 GN FLT4 OR FLT 4.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
 CX NCBI_taxid:10090;
 RN [1]
 RC SEQUENCE FROM N A
 RC STRAIN C57BL/6;
 FX MEDLINE 63348572 PubMed 8403164;
 RA Flannerty H., Kelleher K., Morris G.F., Bean K., Merberg D.M.,
 RA Kriz R., Morris J.C., Stockton H., Turner K.J., Wood C.B.;
 RT Molecular cloning of murine FLT and FLT4;
 RL Oncogene 8:2293-2298(1993)
 CC -1- FUNCTION: RECEPTOR FOR VEGF- α HAS A TYROSINE-PROTEIN KINASE
 CC ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN ADULT LUNG AND LIVER, AND IN
 CC FETAL LIVER, BRAIN, INTESTINE AND PLACENTA.
 CC -1- SIMILARITY: BELONGS TO THE VEGF-1/2/3/4 RECEPTOR FAMILY OF TYROSINE-
 CC PROTEIN KINASES.
 CC -1- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2 TYPE DOMAINS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch)

 EMBL: L07296; AAA40077.1; -
 DR HSSP: P11362; IFGK.
 DR MGI: G95561; Flt4.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR003006; Iq_MHC.
 DR InterPro: IPR003598; Iq_C2.
 DR InterPro: IPR003600; Iq_Like.
 DR InterPro: IPR001824; Receptor_tyr_kin_III.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00047; Iq; 6.
 DR Pfam: PF00069; pkinase; 2.
 DR SMART: SM00410; Ig_Like; 3.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS0107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS00240; RECEPTOR_TYR_KIN_III; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transferase; Tyrosine protein kinase; Phosphorylation; ATP-binding;

KW Receptor; Transmembrane; Signal; Immunoglobulin domain; Repeat;
 glycoprotein.
 FT SIGNAL 1 24 POTENTIAL.
 FT CHAIN 25 1363 VASCULAR ENDOTHELIAL GROWTH FACTOR
 FT DOMAIN 25 775 RECEPTOR 3.
 FT TRANSMEM 776 767 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 798 1363 POTENTIAL.
 FT DOMAIN 1363 1363 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 1363 1363 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 1363 1363 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 1363 1363 IG-LIKE C2-TYPE DOMAIN 3.
 FT DOMAIN 1363 1363 IG-LIKE C2-TYPE DOMAIN 4.
 FT DOMAIN 1363 1363 IG-LIKE C2-TYPE DOMAIN 5.
 FT DOMAIN 1363 1363 IG-LIKE C2-TYPE DOMAIN 6.
 FT DOMAIN 1363 1363 IG-LIKE C2-TYPE DOMAIN 7.
 FT DOMAIN 1363 1363 PROTEIN KINASE.
 FT BINDING 851 859 ATP (BY SIMILARITY).
 FT BINDING 879 879 ATP (BY SIMILARITY).
 FT ACT_SITE 1037 1037 BY SIMILARITY.
 FT DISULFID 51 111 POTENTIAL.
 FT DISULFID 158 206 POTENTIAL.
 FT DISULFID 252 310 POTENTIAL.
 FT DISULFID 445 534 POTENTIAL.
 FT DISULFID 578 653 POTENTIAL.
 FT DISULFID 699 751 POTENTIAL.
 FT MOD_RES 1068 1068 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 33 33 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 104 104 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 166 166 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 251 251 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 299 299 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 411 411 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 515 515 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 527 527 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 582 582 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 594 594 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 683 683 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 690 690 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 758 758 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 1363 AA 153015 MW 153015 MW FLBFA2RCHP99HE9 CR664;

Query Match 69.6% Score 32; DB 1; Length 1363;
 Best Local Similarity 66.7% Pred. No. 110-02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLEPTVEA 9
 |||||
 Db 614 HLEFPLEA 622

Search completed: September 5, 2002, 15:31:42
 Job time: 477 sec


```

QY 1 NLFETPVEA 9
DB 194 NLFETPVEA 202

RESULT 2
ID Q96T85 PRELIMINARY: FFT: 2261 AA.
AC Q96T85;
DT 01-DEC-2001 (Tremblrel, 19, Created)
DT 01-DEC-2001 (Tremblrel, 19, Last sequence update)
DE 01-DEC-2001 (Tremblrel, 19, Last annotation update)
DE ATP-BINDING CASSETTE 1 SUB-FAMILY A MEMBER 1.
GN ABCAL.
OS Homo sapiens (Human).
OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,
OC Mammalia, Eutheria, Primates, Catarrhini, Hominoidea, Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21251004; PubMed=11352667;
RA Qiu Y, Cavalier L, Chiu S, Yand X, Rubin E, Cheng J-F;
RT "Human and Mouse ABCAL Comparative Sequencing and Transgenesis Studies
RT Revealing Novel Regulatory Sequences";
RL Genomics 73:66-76 (2001)
DR EMBL; AF247262; AAK43526.1;
KW ATP binding
SQ SEQUENCE 2261 AA: 254236 MW: 248050408FF041A5 CRC64;

Query Match 82.6%, Score 38, DB 4, Length 2261,
Best Local Similarity 87.5%, Prod. No. 48;
Matches 7, Conservative 1, Mismatches 0, Indels 0, Gaps 0;

QY 1 NLFETPVE 8
DB 808 NLFESPE 815

RESULT 3
Q96S56 PRELIMINARY: PRT: 2261 AA.
ID Q96S56;
AC Q96S56;
DT 01-DEC-2001 (Tremblrel, 19, Created)
DT 01-DEC-2001 (Tremblrel, 19, Last sequence update)
DE 01-DEC-2001 (Tremblrel, 19, Last annotation update)
DE ABCAL.
GN ABCAL.
OS Homo sapiens (Human).
OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,
OC Mammalia, Eutheria, Primates, Catarrhini, Hominoidea, Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tanaka A R, Abe-Dohmae S, Arakawa P, Sadanami K, Kidera A,
RA Kioka N, Amachi T, Yokoyama S, Ikeda K;
RT "A new topological model of functional human ABCAL-Signal peptide-
RT cleavage and glycosylation of a large extracellular domain";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB055982; BAB63210.1;
SQ SEQUENCE 2261 AA: 254400 MW: 212628745242404044;

Query Match 82.6%, Score 28, DB 4, Length 2261,
Best Local Similarity 87.5%, Prod. No. 48;
Matches 7, Conservative 1, Mismatches 0, Indels 0, Gaps 0;

QY 1 NLFETPVE 8
DB 808 NLFESPE 815

RESULT 4

```

```

Q982T3
ID Q982T3 PRELIMINARY: PRT: 393 AA.
AC Q982T3;
DT 01-OCT-2001 (Tremblrel, 18, Created)
DT 01-OCT-2001 (Tremblrel, 18, Last sequence update)
DT 01-OCT-2001 (Tremblrel, 18, Last annotation update)
DE MLR507 PROTEIN.
GN MLR507.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria, Proteobacteria, Alpha subclass, Rhizobiales group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=11092330; PubMed=11214948;
RA Kakeko T, Nakamura Y, Sato S, Asamizu R, Kato T, Sasamoto S,
RA Watanabe A, Idegawa K, Ishikawa A, Kawashima K, Kimura T,
RA Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A,
RA Mochizuki Y, Nakayama S, Nakazaki N, Shimpo S, Sugimoto M,
RA Takeuchi C, Yamada M, Tabata S;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti";
PL DNA Res. 7:331-338 (2000).
DR EMBL; AP004014; BAB54373.1;
DR InterPro; IPR003052; N6_Mtase.
DE PROSITE; PS00692; N6_MTASE; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 393 AA: 43166 MW: 42750338BF59B6B CRC64;

Query Match 80.4%, Score 37, DB 16, Length 393,
Best Local Similarity 77.8%, Prod. No. 12;
Matches 7, Conservative 1, Mismatches 1, Indels 0, Gaps 0;

QY 1 NLFETPVEA 9
DB 173 NLYETPPEA 181

RESULT 5
F78328 PRELIMINARY: PRT: 78 AA.
ID F78328;
AC F78328;
DT 01-MAY-1997 (Tremblrel, 03, Created)
DT 01-MAY-1997 (Tremblrel, 03, Last sequence update)
DT 01-DEC-2001 (Tremblrel, 19, Last annotation update)
DE HP47 PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,
OC Mammalia, Eutheria, Primates, Catarrhini, Hominoidea, Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Ensslin M;
PL Thesis (1996). University of Hannover, Hannover, FRG.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Ensslin M, Vogel T, Calvert J T, Thole H H, Schmidtke J,
RA Matsuda T, Toepfer-Petersen E;
RT "Molecular cloning and characterization of HP47, a novel bear sperm-
RT associated zona pellucida-binding protein homologous to a family of
RT mammalian secretory proteins";
RL Biol. Reprod. 58:1057-1064 (1998).
DR EMBL; Y11718; CAA72405.1;
DR InterPro; IPR000421; FA58_C.
DR PROSITE; PS01286; FA58C_2; 1.
FT NON_TER 1
FT NON_TER 78
SQ SEQUENCE 78 AA: 8696 MW: /E8B04D54b5D4A CRC64;

```

Query Match 78.32, Score 36, FR 4; Length 78;
 Best Local Similarity 75.06; Pred. No. 3;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPVE 8
 |||||
 Db 28 NLFETPVE 35

RESULT 6
 Q9UV31 PRELIMINARY; PRT; 392 AA.
 AC Q9UV31
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 60S RIBOSOMAL PROTEIN L3
 GN RPL3.
 OS Emericella nidulans (Aspergillus nidulans).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocommataceae; Emericella.
 OX NCBI_TaxID=5072;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FGSC4;
 RX MEDLINE-21100459; PubMed-11179686;
 RA Jeong H., Cho G., Han K., Kim J., Min Han D., Jahng K., Chae K.;
 RT "Differential expression of house-keeping genes of Aspergillus
 nidulans during sexual development.",
 PL Gene 267:115-119(2001).
 DP FMRI: AF198447; AAF15509.1;
 DP InterPro: IPR000597; Ribosomal_L3.
 DR Pfam: PF00297; Ribosomal_L3; 1.
 DR ProDom: PD001974; Ribosomal_L3; 1.
 DR PROSITE: PS00474; RIBOSOMAL_L3; UNKNOWN.1.
 KW Ribosomal protein.
 SQ SEQUENCE 392 AA; 44108 MW; 4C20H48O10CFRA4HE CRC64;

Query Match 76.18, Score 35, DB 3; Length 192;
 Best Local Similarity 75.08; Pred. No. 3;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLFETPVE 8
 |||||
 Db 198 NLFETPVE 205

RESULT 7
 Q10295 PRELIMINARY; PRT; 566 AA.
 AC Q10295
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE POLYMERASE.
 GN PLA1.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-96292264; PubMed-8652730.
 RA Onacker M., Minivieille Sebastien L., Keller W.;
 RT "The Schizosaccharomyces pombe pla1 gene encodes a poly(A) polymerase
 and can functionally replace its Saccharomyces cerevisiae homologue.",
 PL Nucleic Acids Res. 24:2585-2591(1996).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN-972H-;
 RA Seeger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
 FL Submitted (AM1999) to the EMBL/GenBank/DBJ databases.
 DE EMBL: X79705; CAA56141.1;
 DE EMBL: AF055216; CAA19081.1;
 DR HSSP: P25500; 1F5A.
 DP InterPro: IPR002934; NTP_transf.
 DR InterPro: IPR001201; PAP_25A_core.
 DR InterPro: IPR000531; TonR_poxC.
 DP Pfam: PF01909; NTP_transf_2; 1.
 DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; UNKNOWN.1.
 SQ SEQUENCE 506 AA; 64108 MW; B8E0E4165AAFD3D5 CRC64;

Query Match 76.18, Score 35; DB 3; Length 566;
 Best Local Similarity 66.78; Pred. No. 47;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLFETPVE 9
 |||||
 Db 36 NLFETPVE 44

RESULT 8
 Q74798 PRELIMINARY; PRT; 658 AA.
 AC Q74798
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE HYPOTHEETICAL 57 kDa PROTEIN C2F10.04 IN CHROMOSOME II.
 GN SFC2D10.04.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA Wood V., Rajandream M.A., Barrell B.G., Taylor K., Harris D.;
 FL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 CC -; SIMILARITY: TO YEAST YJL084L.
 DE EMBL: AL031788; CAA1152.1;
 SQ SEQUENCE 658 AA; 72731 MW; 40EB066BB970B4F7 CRC64;

Query Match 76.18, Score 35; DB 3; Length 658;
 Best Local Similarity 87.58; Pred. No. 56;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLFETPVE 8
 |||||
 Db 423 NLFETPVE 430

RESULT 9
 Q9ZHV4 PRELIMINARY; PRT; 79 AA.
 ID Q9ZHV4
 AC Q9ZHV4;
 DT 01-MAY-1995 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
 DE PEPC (FRAGMENT).
 GN PEPC.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; epsilonbacter group;
 OC Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=F20;
 PX MEDLINE-94454456; PubMed=9780260;

```

RA Hto Y., Arima T., Ito S., Sato H., Miyaji H., Yamazaki Y., Kohli Y.,
RA Kuriyama M.
RT "Full-length sequence analysis of the varA gene from cytotoxic and
RT noncytotoxic Helicobacter pylori";
J. Infect. Dis. 178:1161-1168(1998)
DR EMBL: AF049619; AAD04260.1,
FT NUNLTPR 1
SQ SEQUENCE 79 AA: 9102 MW: 5337PK2167FAVAQF QK*4;

Query Match 73.9%; Score 34; DB 2; Length 79;
Best Local Similarity 75.0%; Pred. No. 91;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LFETPVEA 9
DB 60 LYETPLEA 67
|||||

RESULT 10
Q9SV02 PRELIMINARY: PRT: 177 AA.
AC Q9SV02
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHEtical 20.3 KDA PROTEIN
GN P2206_110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Purnelle B., Boutry M., Goffeau A., Mewes H.W., Ruid S., Lemcke K.,
RA Mayer K.F.X., Quetier F., Salanoubat M.;
RN Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project.
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases
DR EMBL: AL050400; CA443421.1;
FW Hypothetical protein.
SQ SEQUENCE 177 AA: 50206 MW: 84560864AC-RENE QK*4;

Query Match 73.9%; Score 34; DB 10; Length 177;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 NLPETPVEA 9
DB 66 NLPETPVEA 74
|||||

RESULT 11
Q30097 PRELIMINARY: PPT: 261 AA
AC Q30097
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE UBILQUINONE/MENAUQUINONE BIOSYNTHESIS METHYLTRANSFERASE (UBIE).
GN AF0140.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE:98049343; PubMed:9309475;

```

```

RA Klenk H.P., Clayton P.A., Tomb J.F., White O., Nelson V.E.,
RA Ketchum K.A., Dodson P.J., Oxinn M., Hickey P.K., Peterson J.D.,
RA Richardson P.L., Kellavap A.P., Graham D.P., Kyriades N.C.,
RA Fleischman P.D., Quackenbush J., Lee N.H., Sulten G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Eichler C.T., McNeil L.K., Badger J.H., Glisick A., Zhou L.,
RA Overbeek P., Gocayne J.D., Weidman J.F., McDonald L., Orterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine R.P., Sykes S.M.,
RA Sadow P.W., DAndrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
FT "The complete genome sequence of the hyperthermophilic, sulphate-
FT reducing archaeon, Archaeoglobus fulgidus.";
RL Nature 390:584-590(1997).
DR EMBL: AE001096; AAB91083.1;
DR TIGR; AF0140;
DR InterPro: IPR001601; Meth-transf.
DR InterPro: IPR000051; SAM_bind.
DR InterPro: IPR004033; ubiq_C005_methyltransf.
DR Pfam: PF01209; Ubic_methyltran_1.
KW Hypothetical protein; Transferase; Methyltransferase; Ubiquinone;
KW Complete proteome.
SQ SEQUENCE 261 AA: 30256 MW: 6186062208F5B QK*64;

Query Match 73.9%; Score 34; DB 17; Length 261;
Best Local Similarity 75.0%; Pred. No. 33;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLPETPVE 8
DB 35 NLPETPVE 42
|||||

RESULT 12
Q941M0 PRELIMINARY: PPT: 439 AA.
AC Q941M0
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE 2-OXO-3-OXO-3-ARABINO-5-DEPENTEN-1-OXYGENASE
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3712;
RN [1]
RP SEQUENCE FROM N.A.
RA Li G., Quiros C.F.;
RT "Cloning of two major genes involved in aliphatic glucosinolate
RT synthesis in Brassica oleracea.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AY044425; AAK95851.1;
KW Dioxigenase.
SQ SEQUENCE 439 AA: 48313 MW: 87592014AF923F3E QK*64;

Query Match 73.9%; Score 34; DB 10; Length 439;
Best Local Similarity 87.5%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LFETPVEA 9
DB 64 LPETPVEA 71
|||||

RESULT 13
Q9AV92 PRELIMINARY: PRT: 714 AA.
AC Q9AV92
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

```


DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
 DE KINASE-LIKE PROTEIN
 OS Oryza sativa (Rice)
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Liliales; Poaceae;
 CC Erebartoideae; Oryzoideae; Oryza
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. AKITAKOMACHI; TISSUE=LEAF;
 FA Sawada K, Iwata M;
 RT "Oryza sativa kinase like protein."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases
 DR EMBL: AB060276; DAB41205.1;
 DR InterPro: IPR001487; Bromodomain
 DR Pfam: PF03439; Bromodomain_1
 DR PRINTS: PR00503; BROMODOMAIN
 DR SMART: SM00247; HPWC; 1
 DR PROSITE: PS00014; Bromodomain_2; 1
 SQ SEQUENCE 714 AA; 78716 MW; 4P23379CE17C94E9 CF004;

Query Match 73.9%; Score 34; DB 10; Length 714;

Best Local Similarity 55.6%; Pred. No. 10-02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPVEA 9
 :|:|:|:|:|
 DB 185 NIFDPSVDA 193

RESULT 14

QOUTEO PRELIMINARY: PRT; 762 AA.
 AC QOUTEO:
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE HYPOTHETICAL 87.5 KDA PROTEIN Q222.142 IN CHROMOSOME 1.
 GN SPAC222.142.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Hamlin N., Churcher C M., McDougall R.C., Pajandream M.A.,
 RA Bartell B.G.;
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY TO YEAST YOR145W
 DR EMBL: AL132798; CAB60706.1;
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 683 703 POTENTIAL.
 SQ SEQUENCE 762 AA; 87510 MW; 3ADDA2FC8F8E813A CRC64;

Query Match 73.9%; Score 34; DB 3; Length 762;

Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NLFETPVEA 9
 :|:|:|:|:|
 DB 328 NMFEIVIEA 336

RESULT 15

Q983T3 PRELIMINARY: PRT; 54 AA.
 AC Q983T3;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)

DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
 DE MSR8189 PROTEIN.
 GN MSR8189.
 OS Mesophilium loti (Mesophilium loti).
 CC Bacteria; Proteobacteria; Alpha subphylum; Rhizobiaceae group;
 CC Rhizobiaceae; Mesophilium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF403099;
 FA MPELINP-21099930; PubMed-11214968;
 FA Kuno K., Nakamura Y., Sato S., Asanuma E., Kato T., Sasamoto S.,
 FA Watanabe A., Igesawa K., Ishikawa A., Kawashima K., Kimura T.,
 PA Kishida Y., Kiyokawa C., Kohata M., Matsumoto M., Matsuno A.,
 PA Morishiki Y., Nakayama S., Nakaraki N., Shirao S., Sugimoto M.,
 PA Ikeruchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 FI Mesophilium loti."
 FL DNA Pos. 7:331-338(2000).
 DR EMBL: AF004813; RA06477.1;
 KW Complete proteome.
 SQ SEQUENCE 54 AA; 5414 MW; C807E26C4758C859 CRC64;

Query Match 71.7%; Score 33; DB 15; Length 54;

Best Local Similarity 66.7%; Pred. No. 9.9;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 NLFETPVEA 9
 :|:|:|:|:|
 DB 37 DLFRSPVEA 45

Search completed: September 5, 2002, 15:31:04

Job time: 469 sec

XX WPI: 2000-205463/18.
 XX Tumor associated antigen peptides, especially derived from uroplakin,
 PT useful as vaccines to prevent or cure cancers including breast,
 PT bladder, prostate, pancreas, ovary, thyroid, colon and stomach -
 PS Claim 17; Page 100; 113pp; English.
 XX
 CC Tumour associated antigen peptides (TAA) may be used for the
 CC treatment, prevention and cure of cancer or cancer metastases. The
 CC cancer may be breast, bladder, prostate, pancreas, ovary, thyroid,
 CC colon, stomach, head or neck cancer or a carcinoma. The tumour
 CC associated antigens are presentable to the immune system by HLA-A2
 CC molecules and are generally between 8 to 10 amino acids in length.
 CC The amino acids located at positions 2 and 9 of the tumour associated
 CC antigens are the anchor residues which participate in the binding to
 CC MHC class I molecules, more specifically HLA-A2. More tumour
 CC associated antigens are described in the present invention. More tumour
 CC associated antigens are described in records AAY82805-Y82882,
 CC and AAY82855-Y82865 are derived from uroplakin, such as Uroplakin II,
 CC Uroplakin Ia, Uroplakin III and Uroplakin Ib. Those described in
 CC records AAY82825-Y82829 are derived from prostate specific antigen
 CC (PSA). Those described in records AAY82830-Y82835 are derived from
 CC prostate specific membrane antigen (PSMA). Those described in
 CC records Y82836-AAY82839 are derived from prostate acid phosphatase
 CC (PAP). Those described in records AAY82840-Y82846 are derived
 CC from Lactadherin (8A-46). Those described in records AAY82847-Y82854
 CC are derived from Mucin and those described in records AAY82871-Y82882
 CC are derived from Teratocarcinoma derived growth factor (CRIPTO-1).
 XX Sequence 9 AA;

Query Match 100.0%; Score 54; DB 21; Length 9;
 Best local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLOHWVPEL 9
 II IIIIIII
 Db 1 qlqhwvpeL 9
 RESULT 2
 AAY77252
 ID AAY77252 standard; Protein: 387 AA.
 AC AAY77252;
 XX 21-NOV-1995 (first entry)
 DE HMG 46 kDa antigen.
 KW HMG; human milk fat globule; antigen; immunogen; vaccine; virucide;
 KW epithelium; tumor; breast cancer; monoclonal antibody; MA;
 XX Homo sapiens.
 OS Homo sapiens.
 EN W09515171-A.
 DD 08-JUN-1995.
 XX 05-DEC-1994; 94W0-051367.
 XX 03-DEC-1993; 93US-0162402.
 XX (CANC.) CANCER RES FUND CONTRA COSTA.

XX Ceriani RL, Larocca DJ, Peterson JA;
 XX WPI: 1995-215151/28.
 DE N-PSDB: AAG91198.
 XX

PT 46 kD apparent molecular weight human milk fat globule antigen -
 PT used in assays to determine the presence of a cancerous tumour of
 PT epithelial origin, and in a vaccine against neoplastic tumours
 XX Claim 6; Page 46-47; 68pp; English.
 XX
 CC A complete cDNA sequence for the 46 kDa HMG antigen, a major
 CC component of the apical surface of the normal breast epithelial
 CC cell, was obtained by PCR and RACE methods. cDNA clones can be
 CC used to prepare MAbs for use in immunotherapy, immunohistopathology,
 CC prognosis, diagnosis, imaging and therapy. Recombinant antigen can
 CC be expressed in prokaryotic or (glycosylated) in eukaryotic cells.
 XX Sequence 387 AA;

Query Match 100.0%; Score 54; DB 16; Length 387;
 Best local Similarity 100.0%; Pred. No. 0.17;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLOHWVPEL 9
 II IIIIIII
 Db 97 qlqhwvpeL 105
 RESULT 3
 AAY94453
 ID AAY94453 standard; Protein: 387 AA.
 AC AAY94453;
 XX 11-SEP-2000 (first entry)
 DE Human lactadherin protein.
 XX Human lactadherin; MGF-P8; anti-tumour; immune response;
 KW Human; lactadherin; MGF-P8; anti-tumour; immune response;
 KW exosome; dendritic cell.
 XX Homo sapiens.
 OS Homo sapiens.
 PH Key Location/Qualifiers
 FT Peptide 1..23
 FT /label- Secretion_signal
 FT Protein 24..387
 FT /label- Lactadherin
 FT Binding-site 46..48
 FT /label- Integrin_binding_site

XX EP1004664-A1.
 PN 31-MAY-2000.
 XX 24-NOV-1998; 98EP-0402925.
 XX 24-NOV-1998; 98EP-0402925.
 XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX (CURT-) INST CURIE.
 XX WPI: 2000-352567/31.
 DE N-PSDB: AAA27140.
 XX Chimeric isolated (human) lactadherin polypeptide that functions as an
 PT adaptor of cross-priming to eliminate pathogenic antigens, e.g. in
 PT tumors -
 XX Example 3; Page 12; 20pp; English.

XX Lactadherin protein was found in exosomes produced by dendritic cells.
 CC The protein is involved in the phagocytosis of particulate antigens by
 CC dendritic cells. Exosomes produced by dendritic cells exposed to
 CC tumour antigens induce potent immune responses. Lactadherin or variants
 CC of it may be used in the mediation of an immune response. Variants of

CC Lactadherin may be used for inhibition and/or stimulation of the
 CC cross-priming of antigens and stimulation of the phagocytosis of
 CC antigens by dendritic cells. Compositions derived from lactadherin can
 CC also be used to monitor an immune response, more specifically a CTL
 CC (cytotoxic T lymphocyte) response and also to produce cells specific
 CC for a selected antigen. The present sequence is the human
 CC lactadherin protein.
 XX Sequence 387 AA;
 SQ

Query Match 100.0%; Score 54; DB 21; Length 487;
 Best Local Similarity 100.0%; Pred. No. 6.17;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLQHWPEL 9
 Db 97 glqhwpep1105
 |||||
 97 glqhwpep1105

RESULT 4
 AAG75021
 ID AAG75021 standard; Protein: 217 AA.
 AC AAG75021;
 XX 03-SEP-2001 (first entry)
 DT Human colon cancer antigen protein SEQ ID NO:578
 XX Human colon cancer; colon cancer antigen; diagnosis; detection;
 DF colorectal carcinoma
 KW Homo sapiens.
 OS
 XX W020012920-A2.
 PN 05-APR-2001.
 PD 28-SEP-2000; 2000W0-US26524.
 PE 29-SEP-1999; 990S-0157137.
 PR 03-NOV-1999; 990S-0153280.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Ruben SM, Barash SC, Birse CE, Rosen CA,
 PI WPI: 2001-235357/24.
 DP N-PSDB; AAR34426.
 XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
 PT useful for prevention, diagnosing and/or treating colorectal cancers -
 DR Claim 11; Page 7304-7305; 9803pp; English.
 PS AAR32943 to AAR37195 and AAG73514 to AAG77788 represent human colon
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where
 CC the proteins are collectively known as colon cancer antigens. The colon
 CC cancer antigens have cytostatic activity and can be used in gene
 CC therapy and vaccine production. N and P may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate P
 CC expression. For example, N and P may be used to treat disorders
 CC associated with decreased expression by rectifying mutations or deletions
 CC in a patient's genome that affect the activity of P by expressing
 CC inactive proteins or P to supplement the patient's own production of P.
 CC Additionally, N may be used to produce the colon cancer associated P,
 CC by inserting the nucleic acids into a host cell and culturing the cell
 CC to express the proteins. N and P can be used in the prevention, diagnosis
 CC and treatment of colorectal carcinomas and cancers. AAR37196 to AAR37204
 CC and AAR37209 represent sequences used in the exemplification of the
 CC present invention.
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were

CC missing at time of publication, meaning no sequences are present for
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.
 XX
 SQ Sequence 217 AA;
 Query Match 74.1%; Score 40; DB 22; Length 217;
 Best Local Similarity 75.0%; Pred. No. 26;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LOHWPEL 9
 Db 193 lhwpep1200
 |||||
 193 lhwpep1200

RESULT 5
 AAE01423
 ID AAE01423 standard; Protein: 252 AA.
 XX AAE01423;
 AC AAE01423;
 XX 18 JUL 2001 (first entry)
 DT Human secreted protein variant, SEQ ID NO:147.
 DE Human;
 XX Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
 KW foetal abnormality; developmental abnormality; haemopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW parkinson's disease; connective tissue disorder; schizophrenia; skin disorder;
 KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
 KW inflammation; neurological disorder; Alzheimer's disease; food additive;
 KW androgenic disorder; kidney disorder; gastrointestinal disorder; allergy;
 KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
 KW cell culture; chemotaxis; vulnerability; binding partner identification;
 KW gene therapy.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FH Misc-difference 162 /note "Corresponds to any of the naturally occurring
 FT L-amino acids"
 FT Misc-difference 222 /note "Corresponds to any of the naturally occurring
 FT L-amino acids"
 FT Misc-difference 246 /note- "Corresponds to any of the naturally occurring
 FT L-amino acids"
 FT Misc-difference 240 /note- "Corresponds to any of the naturally occurring
 FT L-amino acids"
 FT W0200134629-A1.
 UN 17-MAY-2001.
 XX 08-NOV-2000; 2000W0-US30654.
 XX 12 NOV 1999; 990S 0164845.
 PR 27 JUL 2000; 200003 0221142.
 XX (HUMA-) HUMAN GENOME SCI INC.
 FA Ruben SM, Krasatoulas GA, Wei P, Escobedo M, Baker EP;
 E1 WPI: 2001 329779/24.
 XX New nucleic acid encoding one of 21 human secreted proteins for
 PT diagnosis, prevention, treatment of developmental medical conditions,
 PT such as autoimmune disease and cancer, and used as a food additive or
 PT preservative -
 XX Disclosure; Page 476-477; 490pp; English.
 TS

CC at ftp.wipo.int/pub/published_pt_sequences.

XX Sequence 22 AA;

Query Match: 70.4%; Score 38; DB 22; Length 22;
Best Local Similarity 71.4%; Pred. No. 5.3;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GLOHWP 7
1 |||:
Db 2 qfqlwlp 8

RESULT 10

AAM60107
ID AAM60107 standard; Protein: 22 AA.

XX AC AAM60107;

XX DT 05-NOV-2001 (first entry)

XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 32212.

XX FW Human brain expressed single exon probe expressed in analysis; PT dec.

XX FW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia.

XX FW epilepsy; cancer;

XX OS Homo sapiens.

XX PN W0200157275-A2.

XX PP 09-AUG-2001.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel BK, Chen W, Rank DR;

XX DR WPI: 2001-483446/52.

XX PP 30-JAN-2001; 2001WO-US00667.

XX PP 04-FEB-2000; 2000US-0180312.

XX PP 26-MAY-2000; 2000US-0207456.

XX PP 30-JUN-2000; 2000US-0608408.

XX PP 04-AUG-2000; 2000US-0632466.

XX PP 21-SEP-2000; 2000US-0244887.

XX PP 27-SEP-2000; 2000US-0236359.

XX PP 04-OCT-2000; 2000JP-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel BK, Chen W, Rank DR.

XX DR WPI: 2001-483446/52.

XX PP Single exon nucleic acid probes for analyzing gene expression in human

XX PT brains .

XX PS Example 4; SEQ ID NO: 32212; 650bp . Sequence listing, English.

XX CC The present invention provides a number of single exon nucleic acid

XX CC probes which are derived from genomic sequences expressed in the human

XX CC brain. They can be used to measure gene expression in brain cell samples,

XX CC which may enable the diagnosis and improved treatment of nervous system

XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

XX CC epilepsy and cancers. The present sequence is a protein encoded by one of

XX CC the probes of the invention.

XX SQ Sequence 22 AA;

Query Match

Best Local Similarity 70.4%; Score 38; DB 22; Length 22;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GLOHWP 7
1 |||:
Db 2 qfqlwlp 8

RESULT 11

AAM72718

ID AAM72718 standard; Protein: 22 AA.

XX AC AAM72718;

XX DT 06-NOV-2001 (first entry)

XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 34024.

XX DE Human; bone marrow expressed exon; gene expression analysis; probe;

XX KW microarray; cancer; leukaemia; lymphoma; myeloma.

XX OS Homo sapiens.

XX PN W0200157276-A2.

XX PP 09-AUG-2001.

XX PP 30-JAN-2001; 2001WO-US00668.

XX PP 04-FEB-2000; 2000US-0180312.

XX PP 24-MAY-2000; 2000US-0207456.

XX PP 30-JUN-2000; 2000US-0608408.

XX PP 04-AUG-2000; 2000US-0632466.

XX PP 21-SEP-2000; 2000US-0234687.

XX PP 27-SEP-2000; 2000US-0236359.

XX PP 04-OCT-2000; 2000JP-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel BK, Chen W, Rank DR;

XX DR WPI: 2001-483446/53.

XX PT Human genome-derived single exon nucleic acid probes useful for

XX PP analyzing gene expression in human bone marrow .

XX PS Example 4; SEQ ID NO: 34024; 650bp . Sequence listing, English

XX CC The present invention provides a number of single exon nucleic acid

XX CC probes which are derived from genomic sequences expressed in the human

XX CC bone marrow. They can be used to measure gene expression in bone marrow

XX CC samples, which may enable the improved diagnosis and treatment of cancers

XX CC such as lymphoma, leukaemia and myeloma. The present sequence is a

XX CC protein encoded by one of the probes of the invention.

XX SQ Sequence 22 AA;

Query Match 70.4%; Score 38; DB 22; Length 22;

Best Local Similarity 71.4%; Pred. No. 5.3;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GLOHWP 7

1 |||:
Db 2 qfqlwlp 8

RESULT 12

AAM19701

ID AAM19701 standard; Protein: 22 AA.

XX AC AAM19701;

XX DT 12-OCT-2001 (first entry)

XX DE Peptide #6135 encoded by probe for measuring cervical gene expression.

XX KW Probe; human; microarray; gene expression; cervical epithelial cell;

XX KW cervical cancer.

OS Homo sapiens.
 PN WO200157278 A2.
 XX 04-Aug-2001
 XX 30-JAN-2001; 2001WO-US006670.
 XX 04-PPR-2000; 2000US-0180312.
 PR 25-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-SEP-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0246359.
 PR 04-OCT-2000; 2000US-0242563.
 XX (MOLEF-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR.
 XX WPI; 2001 488901/53
 XX Human genome-derived single-exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells.
 XX Claim 27; SEQ ID NO 24527; 487pp; English.
 XX The present invention relates to human single exon nucleic acid probes
 CC (SENs; see AAI0088 AAI2409). The present sequence is a peptide encoded
 CC by one such probe. The SENs are derived from human Hela cells. The SENs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably
 CC cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX Sequence 22 AA;
 Query Match: 70.4%; Score 38; DB 22; Length 22;
 Best Local Similarity 71.4%; Pred. No. 5.3;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GLQHWVP 7
 Db 1 |||||
 2 gfqhwlp 8
 RESULT 13
 AAM32949
 ID AAM32949 standard, Protein, 22 AA.
 XX AAM32949;
 AC AAM32949;
 DT 17-OCT-2001 (first entry)
 XX Peptide #6986 encoded by probe for measuring placental gene expression.
 DE Probe: microarray, human; placenta, antenatal diagnosis,
 XX genetic disorder.
 KW Homo sapiens.
 OS Homo sapiens.
 XX WO200157272 A2.
 PN 04-Aug-2001
 XX 30-JAN-2001; 2001WO-US006663.

PR 04-PPR-2000; 2000US-0180312.
 PR 25-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-SEP-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0246359.
 PR 04-OCT-2000; 2000US-0242563.
 XX (MOLEF-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR.
 XX WPI; 2001 488901/53.
 XX Human genome-derived single-exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta.
 XX Claim 27; SEQ ID NO 24218; 654pp; English.
 XX The present invention relates to single exon nucleic acid probes (SENP).
 CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.
 XX Sequence 22 AA;
 Query Match: 70.4%; Score 38; DB 22; Length 22;
 Best Local Similarity 71.4%; Pred. No. 5.3;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GLQHWVP 7
 Db 1 |||||
 2 gfqhwlp 8
 RESULT 14
 ABB68244
 ID ABB68244 standard, Protein, 665 AA.
 XX ABB68244;
 AC ABB68244;
 DT 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 31524.
 DE Drosophila, developmental biology; cell signalling, insecticide;
 FW pharmaceutical.
 KW Drosophila melanogaster.
 OS Drosophila melanogaster.
 XX WC200171842 A2
 TN WC200171842 A2
 XX 27-SEP-2001.
 PD 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0241450.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-050800/75.
 DR N PSDB, ABL12347.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell cell
 PT interactions -
 XX

PS Disclosure; SEQ ID NO 31524, 21pp; Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from *Protophila*. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16178-AB16511), expressed tNA
 CC sequences (AB161840-AB16175) and the encoded proteins
 CC (AAH57737-ABH72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 665 AA;

Query Match 70.4%; Score 38; LR 22; Length 665;
 Best local Similarity 71.4%; Pred. No. 1.9e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GLOHWP 7
 |||
 Db 33 qlehwtp 39

RESULT 15

AAE1849
 ID AAE1839 standard; Protein: 4019 AA.

AC AAE1839;

DT 26-FEB-2002 (first entry)

DE Human lung tumour-specific protein SCC2-29.

XX Human; lung tumour protein, immunostimulant; cytostatic; gene therapy;
 KW antisense-therapy; vaccine; immune response; lung cancer, SCC2-29.

XX Homo sapiens.

XX W0200172295-A2.

XX 04-OCT-2001.

XX 28-MAR-2001; 2001WS-0539991.

XX 29-MAR-2000; 2000US-0538037.

XX 05-JUN-2000; 2000US-0588937.

XX 18-AUG-2000; 2000US-0640978.

XX 22-SEP-2000; 2000US 234517P.

XX 01-NOV-2000; 2000US-0704512.

XX 14-DEC-2000; 2000US-0738973.

XX (CORI-) CORIXA CORP.

XX Read SG, Lodes MJ, Mohamath R, Secrist H, Benson DR, Indrias CY;

XX Henderson RA, Fling SP, Algate PA, Elliot M, Mannion J, Kalos MD;

XX WPI; 2001-639201/73.

XX N PSDB; AAD23451.

XX New human lung-specific polynucleotides and polypeptides for the

XX diagnosis and treatment of disease o q lung cancer -

XX Disclosure; Page 309-318; 378pp; English.

XX The invention relates to isolated lung tumour specific proteins and

XX their corresponding cDNA molecules. Lung tumour-specific proteins and

XX their antigen-presenting cells are useful for stimulating and/or

XX expanding T cells specific for a tumour protein, and for inhibiting

XX the development of cancer. The invention also relates to a composition

XX useful for stimulating an immune response, and for treating cancer. The

CC Lung tumour specific oligonucleotides is useful in gene therapy and for
 CC diagnosis, detection and treatment of lung cancer. The present sequence
 CC is human lung tumour-specific protein.
 XX
 SQ Sequence 4019 AA;

Query Match 70.4%; Score 38; LR 22; Length 4019;
 Best local Similarity 85.7%; Pred. No. 1.3e+03;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LQHWPE 8
 |||||
 Db 1475 lqhwpe 1481

Search completed: September 5, 2002, 15:26:31
 Job time: 371 sec



Genome version 4.5
Copyright (c) 1994-2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:28:52, Search time 19.78 seconds
(without alignments)
12.593 Million cell updates/sec

Title: us-09-744-804-39
Perfect score: 54
Sequence: 1 GLOHWVPEL 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DP seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIR71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	77.8	491	2	glycoprotein antiq
2	42	77.8	492	2	pp47 protein pig
3	42	77.8	427	2	fas 6/7 protein pr
4	41	75.9	453	2	rna polymerase - s
5	41	75.9	833	2	atp dependent heli
6	40	74.1	256	2	hypothetical prot
7	40	74.1	285	2	phosphatidylserine
8	40	74.1	453	2	similar to flavoco
9	40	74.1	480	2	hypothetical pro-
10	40	74.1	512	2	probable membrane
11	40	74.1	748	2	hypothetical prote
12	39	72.2	219	2	hypothetical prote
13	39	72.2	238	2	conserved hypoth
14	39	72.2	357	2	hypothetical prote
15	39	72.2	427	2	ads protein precu
16	39	72.2	463	1	milk fat globule m
17	39	72.2	481	2	deoxyribodipyrimid
18	39	72.2	1040	2	probable pnp-fam1
19	39	72.2	1155	2	transcription repa
20	38	70.4	124	2	conserved hypoth
21	38	70.4	238	2	transcription requ
22	38	70.4	331	2	fructose-bisphosph
23	38	70.4	498	2	nucleotide acetylph
24	37	68.5	122	2	hypothetical prote
25	37	68.5	282	2	extended spectrum
26	37	68.5	448	2	hypothetical prote
27	37	68.5	992	2	hypothetical prot-
28	37	68.5	1165	2	mfd protein (AF400
29	37	68.5	1165	2	transcription-topa

30	37	68.5	1170	2	AG3595
31	37	68.5	2143	2	transcription-repa
32	37	68.5	3079	1	hypothetical prote
33	37	68.5	3187	2	prostatic gtpase ac
34	36	66.7	137	2	364k golgi complex
35	36	66.7	221	1	gene hla-bqa2 prot
36	36	66.7	227	2	h-2 class ii histo
37	36	66.7	228	2	mhc class ii histo
38	36	66.7	229	2	mhc class ii histo
39	36	66.7	232	1	mhc class ii histo
40	36	66.7	233	2	ia-alpha polypepte
41	36	66.7	233	2	ia-alpha polypepte
42	36	66.7	243	2	cell surface glyco
43	36	66.7	250	2	class ii histocomp
44	36	66.7	254	1	mhc class ii histo
45	36	66.7	254	1	mhc class ii histo

ALIGNMENTS

RESULT 1
S65138
glycoprotein antiqen M057553, mammary gland, bovine (fragment)
N: Alternate names: glycoprotein component 16/major fat-globule membrane protein/MFG-E
C: Species: Bos primigenius taurus (cattle)
C: Date: 28-Oct-1996 #sequence_revision 15-Mar-1997 #text_change 07-Aug-1998
C: Accession: S65138; G48394
P: Aoki, N., Kishi, M., Taniuchi, Y., Abuchi, T., Nakamura, F., Matsuda, T.
Biochim Biophys Acta 1245, 385-391, 1995
A: Title: Molecular cloning of glycoprotein component 16/globule membrane protein by a cDNA
A: Reference number: S65138; MIM: 6015736
A: Accession: S65138
A: Status: preliminary; not compared with conceptual translation
A: Molecule type: mRNA
A: Residues: 1-401 (AA's)
P: Mathot, I. H.; Bandhelt, L. R.; Lane, W. S.
Biochem Mol Biol Int 29, 545-554, 1993
A: Title: The major fat-globule membrane proteins, bovine components 15/16 and guinea-
II-like sequences.
A: Reference number: A48394; MIM: 93250576
A: Accession: G48394
A: Status: preliminary
A: Molecule type: protein
A: Residues: 267-250 (AA's)
A: Experimental source: milk
A: Note: Sequence extracted from NCBI backbone (NCBI:131457)
C: Species family: milk fat globule protein; disociatin I amine-terminal homology; EGF hom
C: Keywords: glycoprotein
P: 1 32/Domain, EGF domain (fragment) EGF
P: 40 79/Domain, EGF homology, EGF
P: 83-239/Domain, disociatin I amine-terminal homology, IHL
P: 213-401/Domain, disociatin I amine-terminal homology, EGF

Query M197 27,98, Score 45, DB 2, Length 101;
Best local similarity: 77 %, P-val No. 10;
Matches 7, Conservative 0, Mismatches 2, Indels 0, Gaps 0;

QY 1 GLOHWVPEL 9
DB 110 GLOHWVPEL 118

RESULT 2
T11743
pp47 protein - pig (fragment)
C: Species: Sus scrofa domestica (domestic pig)
C: Date: 15 Jul 1999 #sequence_revision 16-Jul-1999 #text_change 21 Jul 2000
C: Accession: T11743
P: Esslin, M., Vogel, T., Gabor, J. J., Thaler, H. H., Schmitz, T.; Matsuda, T.; Top
hol Reprod 58, 1047-1064, 1998
A: Title: Molecular cloning and characterization of pp47, a novel bovine sperm associated

A:Reference number: 217325; MUID:98206817
A:Accession: T11743
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1 409 <ENS>
A:Cross-references: EMBL:Y11683, NID:q2552927, FID:CAA72379.1, FID:q2552928
A:Experimental source: testis
A:Function:
A:Description: may be involved in membrane remodeling and/or function as a DNA polymerase
C:Superfamily: unassigned EGF-related proteins; EGF homology
F:6-40/Domain: EGF homology -EGF

Query Match 77.8%; Score 42; DB 2; Length 409;
Best Local Similarity 77.8%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GLOHWVPEL 9
111 1 111
DB 118 GLOHWVPEL 126

RESULT 3
S74211
PAS-6/7 protein precursor - bovine
C:Species: Bos taurus (cattle)
C:Date: 04-Dec-1997 #sequence_revision 12 Dec 1997 #ext_change 26 Jan 2000
C:Accession: S74211; S78114; S24181
R:Vatregard, J.; Andersen, M.H.; Berglund, L.; Rasmussen, J.T.; Petersen, T.E.
Eur. J. Biochem. 240, 628-636, 1996
A:Title: Characterization of glycoprotein PAS 6/7 from membranes of bovine milk fat globule
A:Reference number: S74211; MUID:97008954
A:Accession: S74211
A:Molecule type: mRNA
A:Residues: 1 423 <NVA>
A:Cross-references: EMBL:X01805; NID:q1632778; FID:CAA62497.1; PID:q1632779
A:Accession: S76114
A:Molecule type: protein
A:Residues: 15 85; 96 119; 146 155; 174 210; 221 242; 248 277; 285 295; 309 337; 350 420; 425 435
R:Kim, D.H.; Kanno, C.; Mizokami, Y.
Biochim. Biophys. Acta 1122, 203-211, 1992
A:Title: Purification and characterization of major glycoproteins, PAS 6 and PAS 7, from
A:Reference number: S23926; MUID:92353107
A:Accession: S24181
A:Molecule type: protein
A:Residues: 363-394 <KIM>
C:Superfamily: milk fat globule protein; discoidin I amino terminal homology; EGF homology
C:Keywords: blocked amino end; disulfide bond; glycoprotein; milk
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-427/Domain: PAS 6; #status #status experimental
F:24-58/Domain: EGF homology <EGF>
F:66-105/Domain: EGF homology <EGF>
F:108-265/Domain: discoidin I amino terminal homology -DBJ
F:249-427/Domain: discoidin I amino terminal homology -DBJ
F:4 479 479-486; 487 490 490-497; 518 518-525; 526 526-533; 534 534-541; 542 542-549; 550 550-557; 558 558-565; 566 566-573; 574 574-581; 582 582-589; 590 590-597; 598 598-605; 606 606-613; 614 614-621; 622 622-629; 630 630-637; 638 638-645; 646 646-653; 654 654-661; 662 662-669; 670 670-677; 678 678-685; 686 686-693; 694 694-701; 702 702-709; 710 710-717; 718 718-725; 726 726-733; 734 734-741; 742 742-749; 750 750-757; 758 758-765; 766 766-773; 774 774-781; 782 782-789; 790 790-797; 798 798-805; 806 806-813; 814 814-821; 822 822-829; 830 830-837; 838 838-845; 846 846-853; 854 854-861; 862 862-869; 870 870-877; 878 878-885; 886 886-893; 894 894-901; 902 902-909; 910 910-917; 918 918-925; 926 926-933; 934 934-941; 942 942-949; 950 950-957; 958 958-965; 966 966-973; 974 974-981; 982 982-989; 990 990-997; 998 998-1005; 1006 1006-1013; 1014 1014-1021; 1022 1022-1029; 1030 1030-1037; 1038 1038-1045; 1046 1046-1053; 1054 1054-1061; 1062 1062-1069; 1070 1070-1077; 1078 1078-1085; 1086 1086-1093; 1094 1094-1101; 1102 1102-1109; 1110 1110-1117; 1118 1118-1125; 1126 1126-1133; 1134 1134-1141; 1142 1142-1149; 1150 1150-1157; 1158 1158-1165; 1166 1166-1173; 1174 1174-1181; 1182 1182-1189; 1190 1190-1197; 1198 1198-1205; 1206 1206-1213; 1214 1214-1221; 1222 1222-1229; 1230 1230-1237; 1238 1238-1245; 1246 1246-1253; 1254 1254-1261; 1262 1262-1269; 1270 1270-1277; 1278 1278-1285; 1286 1286-1293; 1294 1294-1301; 1302 1302-1309; 1310 1310-1317; 1318 1318-1325; 1326 1326-1333; 1334 1334-1341; 1342 1342-1349; 1350 1350-1357; 1358 1358-1365; 1366 1366-1373; 1374 1374-1381; 1382 1382-1389; 1390 1390-1397; 1398 1398-1405; 1406 1406-1413; 1414 1414-1421; 1422 1422-1429; 1430 1430-1437; 1438 1438-1445; 1446 1446-1453; 1454 1454-1461; 1462 1462-1469; 1470 1470-1477; 1478 1478-1485; 1486 1486-1493; 1494 1494-1501; 1502 1502-1509; 1510 1510-1517; 1518 1518-1525; 1526 1526-1533; 1534 1534-1541; 1542 1542-1549; 1550 1550-1557; 1558 1558-1565; 1566 1566-1573; 1574 1574-1581; 1582 1582-1589; 1590 1590-1597; 1598 1598-1605; 1606 1606-1613; 1614 1614-1621; 1622 1622-1629; 1630 1630-1637; 1638 1638-1645; 1646 1646-1653; 1654 1654-1661; 1662 1662-1669; 1670 1670-1677; 1678 1678-1685; 1686 1686-1693; 1694 1694-1701; 1702 1702-1709; 1710 1710-1717; 1718 1718-1725; 1726 1726-1733; 1734 1734-1741; 1742 1742-1749; 1750 1750-1757; 1758 1758-1765; 1766 1766-1773; 1774 1774-1781; 1782 1782-1789; 1790 1790-1797; 1798 1798-1805; 1806 1806-1813; 1814 1814-1821; 1822 1822-1829; 1830 1830-1837; 1838 1838-1845; 1846 1846-1853; 1854 1854-1861; 1862 1862-1869; 1870 1870-1877; 1878 1878-1885; 1886 1886-1893; 1894 1894-1901; 1902 1902-1909; 1910 1910-1917; 1918 1918-1925; 1926 1926-1933; 1934 1934-1941; 1942 1942-1949; 1950 1950-1957; 1958 1958-1965; 1966 1966-1973; 1974 1974-1981; 1982 1982-1989; 1990 1990-1997; 1998 1998-2005; 2006 2006-2013; 2014 2014-2021; 2022 2022-2029; 2030 2030-2037; 2038 2038-2045; 2046 2046-2053; 2054 2054-2061; 2062 2062-2069; 2070 2070-2077; 2078 2078-2085; 2086 2086-2093; 2094 2094-2101; 2102 2102-2109; 2110 2110-2117; 2118 2118-2125; 2126 2126-2133; 2134 2134-2141; 2142 2142-2149; 2150 2150-2157; 2158 2158-2165; 2166 2166-2173; 2174 2174-2181; 2182 2182-2189; 2190 2190-2197; 2198 2198-2205; 2206 2206-2213; 2214 2214-2221; 2222 2222-2229; 2230 2230-2237; 2238 2238-2245; 2246 2246-2253; 2254 2254-2261; 2262 2262-2269; 2270 2270-2277; 2278 2278-2285; 2286 2286-2293; 2294 2294-2301; 2302 2302-2309; 2310 2310-2317; 2318 2318-2325; 2326 2326-2333; 2334 2334-2341; 2342 2342-2349; 2350 2350-2357; 2358 2358-2365; 2366 2366-2373; 2374 2374-2381; 2382 2382-2389; 2390 2390-2397; 2398 2398-2405; 2406 2406-2413; 2414 2414-2421; 2422 2422-2429; 2430 2430-2437; 2438 2438-2445; 2446 2446-2453; 2454 2454-2461; 2462 2462-2469; 2470 2470-2477; 2478 2478-2485; 2486 2486-2493; 2494 2494-2501; 2502 2502-2509; 2510 2510-2517; 2518 2518-2525; 2526 2526-2533; 2534 2534-2541; 2542 2542-2549; 2550 2550-2557; 2558 2558-2565; 2566 2566-2573; 2574 2574-2581; 2582 2582-2589; 2590 2590-2597; 2598 2598-2605; 2606 2606-2613; 2614 2614-2621; 2622 2622-2629; 2630 2630-2637; 2638 2638-2645; 2646 2646-2653; 2654 2654-2661; 2662 2662-2669; 2670 2670-2677; 2678 2678-2685; 2686 2686-2693; 2694 2694-2701; 2702 2702-2709; 2710 2710-2717; 2718 2718-2725; 2726 2726-2733; 2734 2734-2741; 2742 2742-2749; 2750 2750-2757; 2758 2758-2765; 2766 2766-2773; 2774 2774-2781; 2782 2782-2789; 2790 2790-2797; 2798 2798-2805; 2806 2806-2813; 2814 2814-2821; 2822 2822-2829; 2830 2830-2837; 2838 2838-2845; 2846 2846-2853; 2854 2854-2861; 2862 2862-2869; 2870 2870-2877; 2878 2878-2885; 2886 2886-2893; 2894 2894-2901; 2902 2902-2909; 2910 2910-2917; 2918 2918-2925; 2926 2926-2933; 2934 2934-2941; 2942 2942-2949; 2950 2950-2957; 2958 2958-2965; 2966 2966-2973; 2974 2974-2981; 2982 2982-2989; 2990 2990-2997; 2998 2998-3005; 3006 3006-3013; 3014 3014-3021; 3022 3022-3029; 3030 3030-3037; 3038 3038-3045; 3046 3046-3053; 3054 3054-3061; 3062 3062-3069; 3070 3070-3077; 3078 3078-3085; 3086 3086-3093; 3094 3094-3101; 3102 3102-3109; 3110 3110-3117; 3118 3118-3125; 3126 3126-3133; 3134 3134-3141; 3142 3142-3149; 3150 3150-3157; 3158 3158-3165; 3166 3166-3173; 3174 3174-3181; 3182 3182-3189; 3190 3190-3197; 3198 3198-3205; 3206 3206-3213; 3214 3214-3221; 3222 3222-3229; 3230 3230-3237; 3238 3238-3245; 3246 3246-3253; 3254 3254-3261; 3262 3262-3269; 3270 3270-3277; 3278 3278-3285; 3286 3286-3293; 3294 3294-3301; 3302 3302-3309; 3310 3310-3317; 3318 3318-3325; 3326 3326-3333; 3334 3334-3341; 3342 3342-3349; 3350 3350-3357; 3358 3358-3365; 3366 3366-3373; 3374 3374-3381; 3382 3382-3389; 3390 3390-3397; 3398 3398-3405; 3406 3406-3413; 3414 3414-3421; 3422 3422-3429; 3430 3430-3437; 3438 3438-3445; 3446 3446-3453; 3454 3454-3461; 3462 3462-3469; 3470 3470-3477; 3478 3478-3485; 3486 3486-3493; 3494 3494-3501; 3502 3502-3509; 3510 3510-3517; 3518 3518-3525; 3526 3526-3533; 3534 3534-3541; 3542 3542-3549; 3550 3550-3557; 3558 3558-3565; 3566 3566-3573; 3574 3574-3581; 3582 3582-3589; 3590 3590-3597; 3598 3598-3605; 3606 3606-3613; 3614 3614-3621; 3622 3622-3629; 3630 3630-3637; 3638 3638-3645; 3646 3646-3653; 3654 3654-3661; 3662 3662-3669; 3670 3670-3677; 3678 3678-3685; 3686 3686-3693; 3694 3694-3701; 3702 3702-3709; 3710 3710-3717; 3718 3718-3725; 3726 3726-3733; 3734 3734-3741; 3742 3742-3749; 3750 3750-3757; 3758 3758-3765; 3766 3766-3773; 3774 3774-3781; 3782 3782-3789; 3790 3790-3797; 3798 3798-3805; 3806 3806-3813; 3814 3814-3821; 3822 3822-3829; 3830 3830-3837; 3838 3838-3845; 3846 3846-3853; 3854 3854-3861; 3862 3862-3869; 3870 3870-3877; 3878 3878-3885; 3886 3886-3893; 3894 3894-3901; 3902 3902-3909; 3910 3910-3917; 3918 3918-3925; 3926 3926-3933; 3934 3934-3941; 3942 3942-3949; 3950 3950-3957; 3958 3958-3965; 3966 3966-3973; 3974 3974-3981; 3982 3982-3989; 3990 3990-3997; 3998 3998-4005; 4006 4006-4013; 4014 4014-4021; 4022 4022-4029; 4030 4030-4037; 4038 4038-4045; 4046 4046-4053; 4054 4054-4061; 4062 4062-4069; 4070 4070-4077; 4078 4078-4085; 4086 4086-4093; 4094 4094-4101; 4102 4102-4109; 4110 4110-4117; 4118 4118-4125; 4126 4126-4133; 4134 4134-4141; 4142 4142-4149; 4150 4150-4157; 4158 4158-4165; 4166 4166-4173; 4174 4174-4181; 4182 4182-4189; 4190 4190-4197; 4198 4198-4205; 4206 4206-4213; 4214 4214-4221; 4222 4222-4229; 4230 4230-4237; 4238 4238-4245; 4246 4246-4253; 4254 4254-4261; 4262 4262-4269; 4270 4270-4277; 4278 4278-4285; 4286 4286-4293; 4294 4294-4301; 4302 4302-4309; 4310 4310-4317; 4318 4318-4325; 4326 4326-4333; 4334 4334-4341; 4342 4342-4349; 4350 4350-4357; 4358 4358-4365; 4366 4366-4373; 4374 4374-4381; 4382 4382-4389; 4390 4390-4397; 4398 4398-4405; 4406 4406-4413; 4414 4414-4421; 4422 4422-4429; 4430 4430-4437; 4438 4438-4445; 4446 4446-4453; 4454 4454-4461; 4462 4462-4469; 4470 4470-4477; 4478 4478-4485; 4486 4486-4493; 4494 4494-4501; 4502 4502-4509; 4510 4510-4517; 4518 4518-4525; 4526 4526-4533; 4534 4534-4541; 4542 4542-4549; 4550 4550-4557; 4558 4558-4565; 4566 4566-4573; 4574 4574-4581; 4582 4582-4589; 4590 4590-4597; 4598 4598-4605; 4606 4606-4613; 4614 4614-4621; 4622 4622-4629; 4630 4630-4637; 4638 4638-4645; 4646 4646-4653; 4654 4654-4661; 4662 4662-4669; 4670 4670-4677; 4678 4678-4685; 4686 4686-4693; 4694 4694-4701; 4702 4702-4709; 4710 4710-4717; 4718 4718-4725; 4726 4726-4733; 4734 4734-4741; 4742 4742-4749; 4750 4750-4757; 4758 4758-4765; 4766 4766-4773; 4774 4774-4781; 4782 4782-4789; 4790 4790-4797; 4798 4798-4805; 4806 4806-4813; 4814 4814-4821; 4822 4822-4829; 4830 4830-4837; 4838 4838-4845; 4846 4846-4853; 4854 4854-4861; 4862 4862-4869; 4870 4870-4877; 4878 4878-4885; 4886 4886-4893; 4894 4894-4901; 4902 4902-4909; 4910 4910-4917; 4918 4918-4925; 4926 4926-4933; 4934 4934-4941; 4942 4942-4949; 4950 4950-4957; 4958 4958-4965; 4966 4966-4973; 4974 4974-4981; 4982 4982-4989; 4990 4990-4997; 4998 4998-5005; 5006 5006-5013; 5014 5014-5021; 5022 5022-5029; 5030 5030-5037; 5038 5038-5045; 5046 5046-5053; 5054 5054-5061; 5062 5062-5069; 5070 5070-5077; 5078 5078-5085; 5086 5086-5093; 5094 5094-5101; 5102 5102-5109; 5110 5110-5117; 5118 5118-5125; 5126 5126-5133; 5134 5134-5141; 5142 5142-5149; 5150 5150-5157; 5158 5158-5165; 5166 5166-5173; 5174 5174-5181; 5182 5182-5189; 5190 5190-5197; 5198 5198-5205; 5206 5206-5213; 5214 5214-5221; 5222 5222-5229; 5230 5230-5237; 5238 5238-5245; 5246 5246-5253; 5254 5254-5261; 5262 5262-5269; 5270 5270-5277; 5278 5278-5285; 5286 5286-5293; 5294 5294-5301; 5302 5302-5309; 5310 5310-5317; 5318 5318-5325; 5326 5326-5333; 5334 5334-5341; 5342 5342-5349; 5350 5350-5357; 5358 5358-5365; 5366 5366-5373; 5374 5374-5381; 5382 5382-5389; 5390 5390-5397; 5398 5398-5405; 5406 5406-5413; 5414 5414-5421; 5422 5422-5429; 5430 5430-5437; 5438 5438-5445; 5446 5446-5453; 5454 5454-5461; 5462 5462-5469; 5470 5470-5477; 5478 5478-5485; 5486 5486-5493; 5494 5494-5501; 5502 5502-5509; 5510 5510-5517; 5518 5518-5525; 5526 5526-5533; 5534 5534-5541; 5542 5542-5549; 5550 5550-5557; 5558 5558-5565; 5566 5566-5573; 5574 5574-5581; 5582 5582-5589; 5590 5590-5597; 5598 5598-5605; 5606 5606-5613; 5614 5614-5621; 5622 5622-5629; 5630 5630-5637; 5638 5638-5645; 5646 5646-5653; 5654 5654-5661; 5662 5662-5669; 5670 5670-5677; 5678 5678-5685; 5686 5686-5693; 5694 5694-5701; 5702 5702-5709; 5710 5710-5717; 5718 5718-5725; 5726 5726-5733; 5734 5734-5741; 5742 5742-5749; 5750 5750-5757; 5758 5758-5765; 5766 5766-5773; 5774 5774-5781; 5782 5782-5789; 5790 5790-5797; 5798 5798-5805; 5806 5806-5813; 5814 5814-5821; 5822 5822-5829; 5830 5830-5837; 5838 5838-5845; 5846 5846-5853; 5854 5854-5861; 5862 5862-5869; 5870 5870-5877; 5878 5878-5885; 5886 5886-5893; 5894 5894-5901; 5902 5902-5909; 5910 5910-5917; 5918 5918-5925; 5926 5926-5933; 5934 5934-5941; 5942 5942-5949; 5950 5950-5957; 5958 5958-5965; 5966 5966-5973; 5974 5974-5981; 5982 5982-5989; 5990 5990-5997; 5998 5998-6005; 6006 6006-6013; 6014 6014-6021; 6022 6022-6029; 6030 6030-6037; 6038 6038-6045; 6046 6046-6053; 6054 6054-6061; 6062 6062-6069; 6070 6070-6077; 6078 6078-6085; 6086 6086-6093; 6094 6094-6101; 6102 6102-6109; 6110 6110-6117; 6118 6118-6125; 6126 6126-6133; 6134 6134

Db 658 LQHWVPEL 665
|||||

RESULT 6

hypothetical protein 13E11.350 [imported] - Neurospora crassa
C,Species: Neurospora crassa
C,Date: 05-May-2000 #sequence_revision 05 May 2000 #text_change 28-Jul-2000
C,Accession: T48787
R,Schulte, U.; Alim, V.; Habel, J.; Brandt, P.; Fartmann, E.; Holland, F.; Nykattur,
submitted to the Protein Sequence Database, April 2000
A,Reference number: 224541
A,Accession: T48787
A,Status: preliminary
A,Molecule type: DNA
A,Residues: 1-256 <STP>
A,Cross-references: EMBL:AF53420; NDB:00000112; NCSP:13E11.350
A,Experimental source: cosmid contig 13E11; strain 74
C,Genetics:
A,Gene: NCSP:13E11.350
A,Map position: 2
A,Introns: 213/2
C,Superfamily: Neurospora crassa hypothetical protein 13E11.350

Query Match 74.1% Score 40; DB 2; Length 256;
Best Local Similarity 66.7%; Pred. No. 14;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0.

QY 1 GLOHWPEL 9
|||||

Db 141 GLEHWPEV 149

RESULT 7

B82336
phosphatidylserine decarboxylase (P-4 1 1.6%) [similarity] - Vibrio cholerae (strain N
C,Species: Vibrio cholerae
C,Date: 18-Aug-2000 #sequence_revision 20 Aug 2000 #text_change 24-Aug-2001
C,Accession: B82336
R,Heideberg, J.F.; Eisen, T.A.; Nelson, W.C.; Clayton, P.A.; Swinn, M.; Johnson, R.J.;
Richardson, E.; Ermakova, M.D.; Vamathevan, S.; Bass, S.; Qiu, H.; Dragon, L.; Seifert, R.
L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A,Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae
A,Reference number: B82336; MIM:20406833
A,Accession: B82336
A,Status: preliminary
A,Molecule type: DNA
A,Residues: 1-295 <HEH>
A,Cross-references: GB:AF004123; GB:AF004857; NDB:00054745; EMBL:AF004612; GSPHCR0001
A,Experimental source: serogroup O1, strain B16961, biotype El Tor
C,Genetics:
A,Gene: VC0339
A,Map position: 1
C,Superfamily: Escherichia coli phosphatidylserine decarboxylase
C,Keywords: block amino acid; carbon-carbon lyase; carboxyl-lyase
F,251-252/Cleavage site: Gly Ser (autolytic) #status predicted
F,257/Modified site: Iyuvic acid (Ser) (in mature form) #status predicted

Query Match 74.1% Score 40; DB 2; Length 285;
Best Local Similarity 62.5%; Pred. No. 16;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0.

QY 1 GLOHWPEL 8
|||||

Db 7 GLOHWPEV 14

RESULT 8

F96672

Similar to Flavonol 3-O-Glucosyltransferase [imported] - Arabidopsis thaliana
C,Species: Arabidopsis thaliana (mouse ear cress)
C,Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C,Accession: F96672
R,Theologis, A.; Ecker, J.P.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon,
Chiu, C.W.; Chung, M.K.; Conn, A.; Conway, A.R.; Conway, A.P.; Creasy, T.H.; Dewar,
Anson, N.F.; Hughes, P.; Huizar, L.
Nature 408, 816-820, 2000
A,Authors: Hunter, J.; Jenkins, J.; Johnson-Wong, C.; Phan, S.; Kharin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, Z.A.; Luros, J.S.; Maiti, P.; Marzla,
P.T.; M. Pomeroy, T.; Pomeroy, P.; Sakai, H.
A,Authors: Salberg, S.L.; Schwartz, J.P.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo,
Kel, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A,Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A,Reference number: A86141; MIM:21016719
A,Accession: F96672
A,Status: preliminary
A,Molecule type: DNA
A,Residues: 1-452 <STO>
A,Cross-references: GB:AF005174; NDB:00045477; EMBL:AA068000; GSPHCR000141
C,Genetics:
A,Gene: F13011.22
A,Map position: 1

Query Match 74.1% Score 40; DB 2; Length 452;
Best Local Similarity 75.3%; Pred. No. 25;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0.

QY 2 LQHWVPEL 9
|||||

Db 115 LQHWPEM 122

RESULT 9

T10029
hypothetical protein M18628.100 - Mycobacterium leprae
C,Species: Mycobacterium leprae
C,Date: 14-Aug-1999 #sequence_revision 13 Aug 1999 #text_change 21-Jul-2000
C,Accession: T10029
R,Engelmeier, K.; Honore, N.; Woods, S.A.; Caudron, R.; Cole, S.T.
Mol. Microbiol. 7, 197-206, 1993
A,Title: Use of an ordered cosmid library to deduce the genome organization of Mycob
A,Reference number: 216917; MIM:93188700
A,Accession: T10029
A,Status: preliminary; Translated from GB/EMBL/DBJ
A,Molecule type: DNA
A,Residues: 1-480 <SEI>
A,Cross-references: EMBL:Y14957; NDB:02370352; EMBL:CAA75138.1; F15-JC276277
C,Genetics:
A,Note: M18628.100

Query Match 74.1% Score 40; DB 2; Length 489;
Best Local Similarity 85.7%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0.

QY 1 GLOHWPEL 7
|||||

Db 247 GLOHWPEL 253

RESULT 10

G86914
prokalic mactane protein [imported] - Mycobacterium leprae
C,Species: Mycobacterium leprae
C,Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C,Accession: G86914
R,Cole, S.T.; Engelmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.;
R.; Davies, P.M.; Devlin, K.; Duthoy, S.; Fellwell, T.; Fraser, A.; Hamlin, N.; Holmro,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A,Authors: Kottet, S.; Seeger, K.; Simado, S.; Simmonds, M.; Skellern, J.; Squares, R.;

A:Title: Massicot gene decay in the legumes *Bacillus*
 A:Reference number: A66009, MIM:2128262, PMID:11241002
 A:Accession: G86914

A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-512 <STO>
 A:Cross-references: CR A1450380; NID-g13092444; PIDN:CA020555.1; GSPDB:GN00147
 C:Genetics:
 A:Gene: ML0047

Query Match 74 1%; Score 40; PR 2; Length 748;
 Best Local Similarity 85.7%; Pred. No. 29;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GLQHWVP 7
 I I I I I I
 Db 279 GYQHWVP 285

RESULT 11

T30634
 hypothetical protein 322 Mollusum contagiosum virus 1
 N:Alternate names: MC0321
 C:Species: Mollusum contagiosum virus 1
 C:Date: 05-Nov-1999 #sequence_revision 05 Nov-1999 #text_change 21-Jul-2000
 C:Accession: T30634
 R:Senkevich, T.G. Report J.T., Siskind, J.R.; Koenig, F.V.; Parai, G.; Moss, R.
 Science 274: 813-816, 1996
 A:Title: Genome sequence of a human tumorigenic poxvirus: Prediction of specific host re
 A:Reference number: 220876, MIM:36325459
 A:Accession: T30634
 A:Status: preliminary, translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-748 <SEN>
 A:Cross-references: EMBL:060315, NID:g1451343, FIDN:AA055160.1, PID:g1491975
 C:Genetics:
 A:Note: MC0321
 C:Superfamily: vaccinia virus hypothetical protein 49L

Query Match 74 1%; Score 40; PR 2; Length 748;
 Best Local Similarity 85.7%; Pred. No. 43;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 QHWVPEL 9
 I I I I I I
 Db 52 HWVPEL 58

RESULT 12

T37794
 hypothetical protein PA3515 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 03-Dec-1999 #sequence_revision 03 Dec 1999 #text_change 18-Aug-2000
 C:Accession: T37794
 R:Olivier, K.; Harris, D.; Barrell, B.G.; Fajandream, M.A.; Wood, V.
 submitted to the EMBL Data Library, September 1995
 A:Reference number: 221746
 A:Accession: T37794
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-219 <OLI>
 A:Cross-references: EMBL:098529; FIDN:CA011042.1; GSPDB:GN000662; SFDG:SFAC1DE8.14C
 A:Experimental source: strain 972h, cosmid c16E3
 C:Genetics:
 A:Gene: SPDB:SPAC16E8.14C
 A:Map position: 1
 A:introns: 12/2; 32/3; 122/1; 158/3
 C:Superfamily: *Neurospora crassa* hypothetical protein 13E11.350

Query Match 72.2%; Score 39; PR 2; Length 219;

Best Local Similarity 75.0%; Pred. No. 18;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GLQHWVPE 8
 I I I I I I
 Db 110 GLQNWTP 117

RESULT 13

T36025
 conserved hypothetical protein SCC54.12c - Streptomyces coelicolor
 C:Species: *Streptomyces coelicolor*
 C:Date: 03-Dec-1999 #sequence_revision 03 Dec 1999 #text_change 19-May-2000
 C:Accession: T36025
 R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Fajandream, M.A.
 submitted to the EMBL Data Library, March 1999
 A:Reference number: 221581
 A:Accession: T36025
 A:Status: preliminary, translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-338 <SEE>
 A:Cross-references: EMBL:AL035591; PIDN:CA038142.1; GSPDB:GN00070; SCOPDB:SCC54.12c
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOPDB:SCC54.12c
 C:Superfamily: *Bacillus subtilis* conserved hypothetical protein yveC

Query Match 72.2%; Score 39; PR 2; Length 338;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 HWVPEL 9
 I I I I I I
 Db 246 HWVPEL 251

RESULT 14

E83205
 hypothetical protein PA3515 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 18-Sep-2000 #sequence_revision 18-Sep-2000 #text_change 31-Dec-2000
 C:Accession: E83205
 R:Stewart, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.
 ; Lory, S.; Olson, M.V.
 Nature 406: 959-964, 2000
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa
 A:Reference number: AR2450, MIM:20437337
 A:Accession: E83205
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-557 <SL>
 A:Cross-references: CR A650477; CR A650409; NID:g145062; FIDN:AA055403.1; GSPDB:GN
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA3515

Query Match 72.2%; Score 39; PR 2; Length 357;
 Best Local Similarity 22.5%; Pred. No. 29;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LQHWVPEL 9
 I I I I I I
 Db 160 VQHWVPEL 167

RESULT 15

JC4915
 ags protein precursor - rat
 N:Alternate names: O acetyl-Gd3 ganglioside
 C:Species: *Rattus norvegicus* (Norway rat)

C:Date: 26-Sep-1996 #sequence_revision 01 Nov 1996 #text_change 29 Jan 2000
 C:Accession: JC4915
 R:Oura, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanai, Y.
 Biochem. Biophys. Res. Commun. 225, 932-938, 1996
 A:Title: Cloning and expression of cDNA for O-acetylation of GD3 ganglioside.
 A:Reference number: JC4915; M010:96374422
 A:Accession: JC4915
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-427 <GD3>
 A:Cross-references: D05J:064068; N00:q1626956; P00:RAA12210.1; P00:q1626957
 A:Experimental source: CST cell
 C:Comment: This protein is required for the O-acetylation of disialoganglioside sialic acid.
 C:Genetics:
 A:Gene: ags
 C:Superfamily: milk fat globule protein; discoidin I amino-terminal homology: EGF homology
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:28-69/Domain: EGF homology <EG1>
 F:68-107/Domain: EGF homology <EG2>
 F:110-267/Domain: discoidin I amino-terminal homology <DM1>
 F:271-427/Domain: discoidin I amino-terminal homology <DM2>

Query Match 72.2% Score 39; DB 2, length 427;
 Best Local Similarity 77.8%; Pred. No. 36;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GLQHWVPEL 9
 |||||
 Db 138 GLQHWVPEL 146

Search completed: September 5, 2002, 15:28:53
 Job time: 356 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuLink Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:31:42 ; Search time 22.99 seconds
(without alignments)
10.252 Million cell updates/sec

Title: US-09-744-804-39
Perfect score: 54
Sequence: 1 GLQHWVPEL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	Id	Description
1	54	100.0	387	1	MFCM_HUMAN	Q08431 homo sapien
2	42	77.8	409	1	MFCM_PIG	P53485 sus scrofa
3	42	77.8	427	1	MFCM_BOVIN	Q45114 bos taurus
4	39	72.2	219	1	YARE_SCHPO	Q13748 schizosacch
5	39	72.2	427	1	MFCM_PAT	P70492 rattus norv
6	39	72.2	463	1	MFCM_MOUSE	P21956 mus muscula
7	38	70.4	331	1	P16P_PHOSH	P22780 rhodobacter
8	38	70.4	498	1	ACH1_CABEL	P48180 caenorhabdi
9	47	64.5	4074	1	HA21_YEAST	P19158 saccharomyc
10	36	66.7	221	1	HA20_MOUSE	P04227 mus muscula
11	36	66.7	227	1	HA20_MOUSE	P14438 mus muscula
12	36	66.7	232	1	HA24_HUMAN	P04235 homo sapien
13	36	66.7	233	1	HA2F_MOUSE	P14436 mus muscula
14	36	66.7	233	1	HA2P_MOUSE	P14437 mus muscula
15	36	66.7	233	1	HA2S_MOUSE	P14437 mus muscula
16	36	66.7	254	1	HA22_HUMAN	P44254 homo sapien
17	36	66.7	254	1	HA23_HUMAN	P01909 homo sapien
18	36	66.7	254	1	HA27_HUMAN	P05536 homo sapien
19	36	66.7	254	1	HA2C_PIG	P15980 sus scrofa
20	36	66.7	254	1	HA2J_MOUSE	P23150 mus muscula
21	36	66.7	255	1	HA2L_HUMAN	P01408 homo sapien
22	36	66.7	255	1	HA2L_HUMAN	P01407 homo sapien
23	36	66.7	255	1	HA2E_HUMAN	P01406 homo sapien
24	36	66.7	255	1	HA2B_PIG	P15781 sus scrofa
25	36	66.7	256	1	HA2B_MOUSE	P14434 mus muscula
26	36	66.7	256	1	HA2B_RAT	P20037 rattus norv
27	36	66.7	256	1	HA2D_MOUSE	P04228 mus muscula
28	36	66.7	256	1	HA2C_MOUSE	P01916 mus muscula
29	36	66.7	529	1	CN1A_BOVIN	P14103 bos taurus
30	36	66.7	534	1	CN1A_HUMAN	P54750 homo sapien
31	36	66.7	557	1	CN1A_NEDCR	P03945 neotospira
32	36	66.7	565	1	CN1A_MOUSE	Q614W1 mus muscula
33	36	66.7	1101	1	YA4G_SCHPO	Q09733 schizosacch

ALIGNMENTS

RESULT 1

MFCM_HUMAN

ID MFCM_HUMAN STANDARD; PRI: 487 AA.

AC Q08431;

DI 01-OCT-1996 (rel. 34, created)

DE 01-NOV-1997 (rel. 35, last sequence update)

DE 16-OCT-2001 (rel. 40, last annotation update)

DE Lactadherin precursor (Milk fat globule EGF factor 8) (MFG-EB) (HMF6)

DE (Breast epithelial antigen BA46) (MFCM) [Contains: Modin].

GN MFCP8.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP TISSUE=Breast, and Breast carcinoma;

EX MEDLINE-9621903; PubMed-8639264;

RA Couto J P, Taylor M P, Cowlin S G, Coriani P L, Peterson J A,;

RT "Cloning and sequence analysis of human breast epithelial antigen

BA46 reveals an EGF cell adhesion sequence presented on an epidermal

growth factor-like domain.";

RL DNA Cell Biol. 15:281-286(1996).

RN [2]

RP TISSUE=Mammary gland;

EX MEDLINE-9337151; PubMed-1305932;

EA Lefevre B, Peterson J A, Uffee K, Kato Yoshi J, Bistrain A M,;

RA Coriani P L,;

RT "A Mr 46,000 human milk fat globule protein that is highly expressed

in human breast tumors contains factor VIII-like domains.";

EL Cancer Res. 51:494-498(1991).

RN [3]

RP PARTIAL SEQUENCE, AND CHARACTERIZATION.

EX MEDLINE-9819424; PubMed-9530276;

EA Giordano M G, Giacchino M, Giunta G, Genti A,;

RA Giordano M,;

RT "Isolation and characterization of full and truncated forms of human

breast carcinoma protein BA46 from human milk fat globule membranes.";

EL J. Protein Chem. 17:143-148(1998).

RN [4]

RP SEQUENCE OF 268-317, AND IDENTIFICATION OF MEDIN.

EX MEDLINE-96342076; PubMed-10411933;

EA Buckdahl B, Roslund J, Stetten K, Westmark G, Westmark G,;

RA Ljundberg L G, Nordstam C, Enderstrom U, Westmark G,;

RT "Medin, an integral fragment of active smooth muscle cell-produced

Lactadherin forms the most common human amyloid.";

EL Proc. Natl. Acad. Sci. U.S.A. 96:8669-8674(1999).

RN [5]

RP CHARACTERIZATION.

EX MEDLINE-9745846; PubMed-9260029;

RA Taylor M P, Couto J P, Scallan C D, Coriani P L, Peterson J A,;

RT "Lactadherin (formerly BA46), a membrane associated glycoprotein

expressed in human milk and breast carcinomas, promotes Atp Gly-Asp

(RGD) dependent cell adhesion.";

RL DNA Cell Biol. 16:861-869(1997).
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
 CC SPECIFICALLY TO PTAVIANS AND INHIBITS ITS REPLICATION.
 CC -1- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID.
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
 CC MEDIA. OVEREXPRESSED IN SEVERAL CARCINOMAS.
 CC -1- PIM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
 CC AMINO ACID 264 AND 273.
 CC -1- SIMILARITY: CONTAINS 2 EGF LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

EMBL: U58516; AAC50549.1; -
 EMBL: S56151; AAB1971.1; -
 MIM: 602281; -

InterPro: IPR000561; EGF-like
 InterPro: IPR000421; FA58_C.
 Pfam: PF00008; EGF; 1.
 Pfam: PF00754; F5/P8_Type_C; 2.
 SMART: SM00181; EGF; 1.
 SMART: SM00241; FASG; 2.
 PROSITE: PS00022; EGF_1; 1.
 PROSITE: PS01186; EGF_2; 1.
 PROSITE: PS01285; FASG; 1; 2.
 PROSITE: PS01285; FASG; 1; 2.
 PROSITE: PS01286; FASG; 2; 2.
 Signal: Glycoprotein-Milk, Repeat, EGF-like domain; Amyloid.

FT SIGNAL 1 24
 FT CHAIN 24 387 LACTADHERIN.
 FT CHAIN 202 387 LACTADHERIN; SUGT FORM.
 FT CHAIN 268 317 MEDIN.
 FT DOMAIN 24 67 EGF-LIKE.
 FT DOMAIN 70 225 F5/8 TYPE C 1.
 FT DOMAIN 240 387 F5/8 TYPE C 2.
 FT SITE 46 48 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 27 38 BY SIMILARITY.
 FT DISULFID 32 55 BY SIMILARITY.
 FT DISULFID 57 66 BY SIMILARITY.
 FT DISULFID 70 225 BY SIMILARITY.
 FT DISULFID 212 216 BY SIMILARITY.
 FT DISULFID 230 387 BY SIMILARITY.
 FT CARBOHYD 238 238 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 325 325 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 329 329 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 350 350 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 387 AA: 4123 MW: 26857.160x78.30x664.

Query Match 100.0%; Score 54; DH 1; Length 387;
 Best Local Similarity 100.0%; Pred. No. 0.019;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLOHWVPEL 9
 |||||
 Db 97 GLOHWVPEL 105

RESULT 2
 MEGM_PIG STANDARD; PKT; 409 AA.
 AC P79385;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 04-MAY-2000 (Rel. 39, Last annotation update)
 DE Lactadherin (Milk fat globule-EGF factor 8) (MEG-F8) (MEGM) (Sperm
 surface protein SP47) (p047).

GN MFCF8
 OS Sus scrofa (pig).
 CC Euryota: Mammalia: Chordata: Craniata: Vertebrata: Euteleostomi:
 CC Mammalia: Eutheria: Cetartiodactyla: Suina: Suidae: Sus.
 CC NCBI_TaxID=9823;
 RN 11
 RP SEQUENCE FROM N.A.
 RC TISSUE-Testis;
 RA Easlin M.A.;
 FC Submitted (NCV-1497) to the EMBL/GenBank/CCO databases.
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
 CC BINDING PROTEIN.
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
 CC SPERMATOZOAN.
 CC -1- SIMILARITY: CONTAINS 2 EGF LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

EMBL: Y11683; CAA72379.1; -
 HSSP: P00740; IEDM.
 InterPro: IPR000561; EGF-like.
 InterPro: IPR000421; FA58_C.
 Pfam: PF00008; EGF; 2.
 Pfam: PF00754; F5/P8_Type_C; 2.
 SMART: SM00191; EGF; 2.
 SMART: SM00241; FASG; 2.
 PROSITE: PS00022; EGF_1; 2.
 PROSITE: PS01186; EGF_2; 2.
 PROSITE: PS01285; FASG; 1; 2.
 PROSITE: PS01286; FASG; 2; 2.
 Glycoprotein, Repeat, EGF like domain.

FT DOMAIN 2 41 EGF-LIKE 1.
 FT DOMAIN 44 88 EGF-LIKE 2.
 FT DOMAIN 91 247 F5/8 TYPE C 1.
 FT DOMAIN 252 409 F5/8 TYPE C 2.
 FT SITE 67 69 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 6 17 BY SIMILARITY.
 FT DISULFID 11 29 BY SIMILARITY.
 FT DISULFID 31 40 BY SIMILARITY.
 FT DISULFID 91 247 BY SIMILARITY.
 FT DISULFID 234 238 BY SIMILARITY.
 FT DISULFID 252 409 BY SIMILARITY.
 FT CARBOHYD 41 41 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 372 372 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 409 AA: 4572 MW: 100741.60x259.27x664.

Query Match 77.8%; Score 42; DH 1; Length 409;
 Best Local Similarity 77.8%; Pred. No. 2.8;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GLOHWVPEL 9
 |||||
 Db 118 GLOHWVPEL 126

RESULT 3
 MEGM_BOVIN STANDARD; PKT; 427 AA.
 AC Q95114; Q27959; P79344;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-2001 (Rel. 40, Last annotation update)
 DE Lactadherin (Milk fat globule-EGF factor 8) (MEG-F8)
 DE (M35753); (FA5-5/FA5-7 glycoprotein) (MEGM) (Sperm surface protein

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL organisation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announcements/](http://www.isb-sib.ch/announcements) or send an email to license@sib.ski.ch.)

RN SEQUENCE FROM N.A.
 RP MEDLINE-83285339; PubMed-6309407;
 RA Benoit C.O., Mathis D.J., Kanter M.R., Williams V.E., McDewitt H.O.;
 RT "Regions of allelic hypervariability in the murine A alpha immune
 response gene".
 RI Cell 34:169-177(1983).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch>)
 or send an email to license@sib-sib.ch.
 CC
 DR EMEL; K01925; AAA79623.1; -;
 DR HSSP; P01910; IIAK.
 DR InterPro; IPR004006; Iq_MHC.
 DR InterPro; IPR00407; Iq_C1.
 DR InterPro; IPR001003; MHC_II_alpha.
 DR Pfam; PF00047; Iq; 1.
 DR Pfam; PF00993; MHC_II_alpha; 1.
 DR SMART; SM00407; IqC1; 1.
 DR PROSITE; PS00290; Iq_MHC; 1.
 DR MHC II; Transmembrane; Glycoprotein.
 KW NON_TER 1
 FT DOMAIN <1 76 EXTRACELLULAR ALPHA-1.
 FT DOMAIN 77 170 EXTRACELLULAR ALPHA-2.
 FT DOMAIN 171 183 CONNECTING PEPTIDE.
 FT TRANSMEM 184 209
 FT DOMAIN 210 221 CYTOPLASMIC TAIL.
 FT DISULFID 99 155 BY SIMILARITY.
 FT CARBOHYD 110 110 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 221 AA; 24461 MW; 50016FAC3729151; CPC64;
 Query Match 66.7%; Score 25; DR 1; Length 241;
 Best Local Similarity 62.5%; Pred. No. 17;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LOHWVPEL 9
 DB 167 LKHWEPEL 174
 RESULT 11
 ID HAZV_MOUSE STANAEQ. PRT. 227 AA.
 AC P14438;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DE H-2 class II histocompatibility antigen, A alpha chain (fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Benoit C.O., Mathis D.J., Kanter M.R., Williams V.E., McDewitt H.O.;
 RT "Regions of allelic hypervariability in the murine A alpha immune
 response gene".
 RI Cell 34:169-177(1983).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch>)
 or send an email to license@sib-sib.ch.
 CC
 DR EMEL; K01925; AAA79623.1; -;
 DR HSSP; P01910; IIAK.
 DR InterPro; IPR004006; Iq_MHC.
 DR InterPro; IPR00407; Iq_C1.
 DR InterPro; IPR001003; MHC_II_alpha.
 DR Pfam; PF00047; Iq; 1.
 DR Pfam; PF00993; MHC_II_alpha; 1.
 DR SMART; SM00407; IqC1; 1.
 DR PROSITE; PS00290; Iq_MHC; 1.
 DR MHC II; Transmembrane; Glycoprotein.
 KW NON_TER 1
 FT DOMAIN <1 76 EXTRACELLULAR ALPHA-1.
 FT DOMAIN 77 170 EXTRACELLULAR ALPHA-2.
 FT DOMAIN 171 183 CONNECTING PEPTIDE.
 FT TRANSMEM 184 209
 FT DOMAIN 210 221 CYTOPLASMIC TAIL.
 FT DISULFID 99 155 BY SIMILARITY.
 FT CARBOHYD 110 110 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 221 AA; 24461 MW; 50016FAC3729151; CPC64;
 Query Match 66.7%; Score 25; DR 1; Length 241;
 Best Local Similarity 62.5%; Pred. No. 17;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LOHWVPEL 9
 DB 167 LKHWEPEL 174
 RESULT 12
 ID HAZ4_HUMAN STANDARD; PRT. 232 AA.
 AC P04225;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 20-MAR-1987 (Rel. 04, Last sequence update)
 DE H-2 class II histocompatibility antigen, E alpha chain precursor
 DE HLA class II histocompatibility antigen, E alpha chain precursor
 DE (DQ-B*09 alpha chain) (fragment).
 GN HLA-DQA1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-95216512; PubMed-3879967;
 RA Mitsuuchi T., Mitsuuchi T., Silver J.;
 RT "Nucleotide sequence of an HLA-DQ alpha chain derived from a DRW9
 cell line: genetic and evolutionary implications".
 RI Proc. Natl. Acad. Sci. U.S.A. 82:3420-3424(1985).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch>)
 or send an email to license@sib-sib.ch.
 CC
 DR EMEL; M11124; AAA59754.1; -;
 DR PIP; A02213; HIRH9Q.
 DR HSSP; P01910; IIAK.
 DR MIM; 146880; -;
 DR InterPro; IPR004006; Iq_MHC.
 DR InterPro; IPR004597; Iq_C1.
 DR InterPro; IPR001003; MHC_II_alpha.
 DR Pfam; PF00047; Iq; 1.
 DR Pfam; PF00993; MHC_II_alpha; 1.
 DR SMART; SM00407; IqC1; 1.
 DR PROSITE; PS00290; Iq_MHC; 1.
 DR MHC II; Transmembrane; Glycoprotein.

CC or send an email to license@sib-sib.ch.

CC
 DR EMEL; K01925; AAA79623.1; -;
 DR HSSP; P01910; IIAK.
 DR InterPro; IPR004006; Iq_MHC.
 DR InterPro; IPR00407; Iq_C1.
 DR InterPro; IPR001003; MHC_II_alpha.
 DR Pfam; PF00047; Iq; 1.
 DR Pfam; PF00993; MHC_II_alpha; 1.
 DR SMART; SM00407; IqC1; 1.
 DR PROSITE; PS00290; Iq_MHC; 1.
 DR MHC II; Transmembrane; Glycoprotein.
 KW NON_TER 1
 FT DOMAIN <1 82 EXTRACELLULAR ALPHA-1.
 FT DOMAIN 83 176 EXTRACELLULAR ALPHA-2.
 FT DOMAIN 177 189 CONNECTING PEPTIDE.
 FT TRANSMEM 190 215
 FT DOMAIN 216 227 CYTOPLASMIC TAIL.
 FT DISULFID 105 161 BY SIMILARITY.
 FT CARBOHYD 116 116 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 227 AA; 25115 MW; 0167B9F10006977; CPC64;
 Query Match 66.7%; Score 36; DR 1; Length 227;
 Best Local Similarity 62.5%; Pred. No. 18;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LOHWVPEL 9
 DB 173 LKHWEPEL 180
 RESULT 12
 ID HAZ4_HUMAN STANDARD; PRT. 232 AA.
 AC P04225;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 20-MAR-1987 (Rel. 04, Last sequence update)
 DE H-2 class II histocompatibility antigen, E alpha chain precursor
 DE HLA class II histocompatibility antigen, E alpha chain precursor
 DE (DQ-B*09 alpha chain) (fragment).
 GN HLA-DQA1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-95216512; PubMed-3879967;
 RA Mitsuuchi T., Mitsuuchi T., Silver J.;
 RT "Nucleotide sequence of an HLA-DQ alpha chain derived from a DRW9
 cell line: genetic and evolutionary implications".
 RI Proc. Natl. Acad. Sci. U.S.A. 82:3420-3424(1985).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch>)
 or send an email to license@sib-sib.ch.
 CC
 DR EMEL; M11124; AAA59754.1; -;
 DR PIP; A02213; HIRH9Q.
 DR HSSP; P01910; IIAK.
 DR MIM; 146880; -;
 DR InterPro; IPR004006; Iq_MHC.
 DR InterPro; IPR004597; Iq_C1.
 DR InterPro; IPR001003; MHC_II_alpha.
 DR Pfam; PF00047; Iq; 1.
 DR Pfam; PF00993; MHC_II_alpha; 1.
 DR SMART; SM00407; IqC1; 1.
 DR PROSITE; PS00290; Iq_MHC; 1.
 DR MHC II; Transmembrane; Glycoprotein.

DT 01-JAN-1990 (Rel. 13, last annotation update)
 DE H-2 class II histocompatibility antigen, A-S alpha chain.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85190610; PubMed 2581258;
 RA Landais D., Matthes H., Hemoist C., Mathis D.;
 RT "A molecular basis for the Ia.2 and Ia.19 antigenic determinants";
 RC Proc. Natl. Acad. Sci. U.S.A. 82:2940-2944(1985).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M11358; AAA39622.1; -;
 DR HSSP: P01910; IIAK.
 DR InterPro: IPR003006; Iq_MHC.
 DR InterPro: IPR003597; Iq_c1.
 DR InterPro: IPR001003; MHC_II_alpha.
 DR Pfam: PF00047; Iq; 1.
 DR Pfam: PF00993; MHC_II_alpha; 1.
 DR SMART: SM00407; IGc1; 1.
 DR PROSITE: PS00290; IG_MHC; 1.
 KW MHC II; Transmembrane; Glycoprotein.
 FT DOMAIN 1 88
 FT DOMAIN 89 182 EXTRACELLULAR ALPHA-1.
 FT DOMAIN 183 195 EXTRACELLULAR ALPHA-2.
 FT TRANSMEM 196 221 CONNECTING PEPTIDE.
 FT DOMAIN 222 233
 FT DOMAIN 222 233 CYTOPLASMIC TAIL.
 FT DISULFID 111 167 BY SIMILARITY.
 FT CARGOXYD 122 123 N-TERMINAL GLYCAN (GLYCAN).
 SQ SEQUENCE 233 AA; 25801 MW; 6652ED01A87F3019 CRC64;

Query Match 66.7%; Score 36; DB 1; Length 233;
 Best Local Similarity 62.5%; Pred. No. 18;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LQHWVPEL 9
 Db 179 LKHWPEI 186

Search completed: September 5, 2002, 15:31:42
 Job time: 477 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:31:04 : Search time 170.86 Seconds
(without alignments)
12.673 Million cell updates/sec

Title: US-09-744-804-39
Perfect score: 54
Sequence: 1 QHWHVPEL 9

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 56222 seqs, 172934329 residues
Total number of hits satisfying chosen parameters: 56222

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL19:
1: sp.archaea:
2: sp.bacteria:
3: sp.fungi:
4: sp.human:
5: sp.invertebrate:
6: sp.mammal:
7: sp.mhc:
8: sp.organelle:
9: sp.phage:
10: sp.plant:
11: sp.rodent:
12: sp.virus:
13: sp.vertebrate:
14: sp.unclassified:
15: sp.virus:
16: sp.bacteriaph:
17: sp.archaeo:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	100.0	335	4	Q9BTL9 homo sapien
2	46	85.2	353	6	Q77718 equus caball
3	41	75.9	453	16	P77967 synochocyst
4	41	75.9	521	3	Q93654 vibrio sp
5	41	75.9	833	16	Q93654 vibrio sp
6	40	74.1	256	3	Q93654 vibrio sp
7	40	74.1	285	16	Q93654 vibrio sp
8	40	74.1	337	12	Q11317
9	40	74.1	452	10	Q93654
10	40	74.1	486	3	Q11317
11	40	74.1	512	16	Q93654
12	40	74.1	748	12	Q93654
13	40	74.1	1605	2	Q93654
14	39	72.2	61	10	Q93654
15	39	72.2	66	5	Q93654
16	39	72.2	183	8	Q93654

17 39 72.2 192 11 Q9CWO1
18 39 72.2 210 1 Q977V4
19 39 72.2 338 2 Q9513
20 39 72.2 357 16 Q9HY94
21 39 72.2 425 11 Q9WIS3
22 39 72.2 463 11 Q9BIX9
23 39 72.2 481 16 Q9HVD2
24 39 72.2 1155 16 Q9A782
25 39 72.2 159 16 Q999Y1
26 38 70.4 183 8 Q9XND2
27 38 70.4 238 16 Q9A510
28 38 70.4 244 16 Q9A510
29 38 70.4 635 10 Q9A510
30 38 70.4 655 5 Q9W241
31 38 70.4 740 3 Q9A154
32 38 70.4 1335 4 Q95222
33 38 70.4 4025 4 Q9NR13
34 37 68.5 67 10 Q42070
35 37 68.5 117 10 Q9HWC
36 37 68.5 122 8 Q95168
37 37 68.5 134 7 Q95548
38 37 68.5 158 13 Q9A475
39 37 68.5 158 13 Q9A475
40 37 68.5 183 8 Q9XN08
41 37 68.5 223 4 Q9H0G3
42 37 68.5 240 1 Q9A510
43 37 68.5 274 4 Q95222
44 37 68.5 286 2 Q46954
45 37 68.5 286 2 Q46954

ALIGNMENTS

RESULT 1

Q9BTL9 PRELIMINARY; PRT; 335 AA.
AC Q9BTL9; 01-JUN-2001 (TREMBL19, 17, Created)
DT 01-JUN-2001 (TREMBL19, 17, Last sequence update)
DT 01-DEC-2001 (TREMBL19, 19, Last annotation update)
DE SIMILAR TO MILK FAT GLOBULE-EGF FACTOR 8 PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cladista; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
PC TISSUE-MELANOMA;
FA Strausberg P.
KL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -15. SIMILARITY: CONTAINS 1 F5/6 TYPE C DOMAIN.
DR EMBL; BC003610; AAH03610.1; -.
DR HSSP; P08709; 1BF9.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001438; EGF-1.
DR InterPro; IPR000421; FAS8_C.
DR PRINTS; R00010; EGFPROL.
DR SMART; SM00181; EGF, 1.
DR SMART; SM00301; EGF-like, 1.
DR SMART; SM00301; FAS8_C, 2.
DR EASite; F50126; EGF-1; UERKMR_1.
DR EASite; F50126; EGF-1; UERKMR_1.
DR PROSITE; PS01289; FAS8C_1; 1.
DR PROSITE; PS01289; FAS8C_1; 1.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 335 AA, 37523 MW, 20884 EMBL78E9E25 Q9C64;

Query Match 100.0%; Score 54; DB 4; Length 335;
Best local similarity 100.0%; Pred. No. 9.12;
Matches 3; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GLQHWVPEL 9
DB 97 GLQHWVPEL 105

RESULT 2
077718 077718 PPF:IMINAPY; PPT; 364 AA
AC 077718;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SPERM-MEMBRANE ASSOCIATED PROTEIN P47 (EPAGMENT).
OS Equus caballus (Horse).
OC Eukaryota, Metazoa, Chordata, Cladacea, Vertebrata; Euteleostomi,
OC Mammalia, Eutheria; Perissodactyla; Equidae; Equus.
CX NCBI_TaxID=9796;
RN [1]
RP TISSUE:TESTIS;
RC TISSUE:TESTIS;
RA Gentzel M., Toepfer-Petersen E.;
KL Submitted (AUG-1998) to the EMBL/GenBank/TrEMBL databases
DR EMBL: AJ010121; CAA00010;
DR HSSP: P00740; IEDM.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000421; FA58_1.
DR Pfam: PF00008; EGF; 1.
DR Pfam: PF00754; F5_E8_1; YP_5; 2.
DR SMART: SM00181; EGF; 1.
DR SMART: SM00231; FA58C; 2.
DR PROSITE: PS00022; EGF_1; UNP:WN_2
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS01285; FA58C; 1; 1.
KW EGF-like domain; Glycoprotein.
FT NON_TER 1 363
FT NON_TER 363 363
SQ SEQUENCE 363 AA: 46744 MW: 14865355APK23340 P00754;

Query Match 85.2%; Score 45; DB 6; Length 363;
Best Local Similarity 88.9%; Pred. No. 3.1;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GLQHWVPEL 9
DB 102 GLQHWVPEL 110

RESULT 3
P77967 077967 PRELIMINARY; PPT; 453 AA.
AC P77967;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE DNA PHOTOLYASE.
GN PHR OR SLL1629.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
CX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuroda T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,
RA Hosonishi T., Matsuno A., Muraki A., Nakazaki N., Naraoka F., Okamura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
KT Synchocystis sp strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein coding regions."
RL DNA Res. 3:109-136 (1996).
DR EMBL: D00000; BAA17766.1;

```

```

DR HSSP: P05327; IONP.
DR InterPro: IPR000474; DNA_photolyase.
DR InterPro: IPR002381; DNA_photolyase_1.
DR Pfam: PF00875; DNA_photolyase; 1.
DR ProDom: PD004390; DNA_photolyase; 1.
DR PROSITE: PS00394; DNA_PHOTOLYASES_1_1; 1.
KW Lyase; Complete proteome.
SQ SEQUENCE 453 AA: 52920 MW: 58555000G5E5K2 C0C64;

Query Match 75.9%; Score 41; DB 16; Length 453;
Best Local Similarity 75.0%; Pred. No. 29;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LQHWVIEL 9
DB 392 LPHWPEL 399

RESULT 4
093P54 093P54 PPF:IMINAPY; PPT; 521 AA.
AC 093P54;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ALKALINE PHOSPHATASE.
OS Vibrio sp. G15-21.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
CX NCBI_TaxID=169049;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G15-21;
RA Asgeirsson B., Andresson O.S.;
RT Sp., as deduced from the nucleotide sequence."
PI Biochim Biophys Acta 990-6(2001).
DR EMBL: AF352014; AAK94204.1; -.
SQ SEQUENCE 521 AA: 57384 MW: 60649304D06CF102 C0C64;

Query Match 75.9%; Score 41; DB 2; Length 521;
Best Local Similarity 62.5%; Pred. No. 33;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLQHWVPE 8
DB 170 GLRHWIPK 177

RESULT 5
Q9PDZ9 09PDZ9 PPF:IMINAPY; PPT; 833 AA.
AC Q9PDZ9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ATP-DEPENDENT HELICASE.
CN XFI229.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
CX Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
PC STEAIN-9A5C;
FX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.T.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarado P., Alves I.M.C., Azevedo J.P., Bello J.M., Briones M.R.S.,
RA Barros M.H., Benavides E.B., Bordin S., Bove J.M., Briones M.R.S.,
RA Borge M.H.P., Camargo A., Camargo L.F., Carraro D.M., Carter H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto F., Deena C., El Dory H.,

```

RA Faciniani A P., Ferreira A.J.S., Ferreira V.C.A., Fetto J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohne M., Furian L.P.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hobeisel T.D., Jungheirna M.T., Kemper F.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.F., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos F.G.M., Lemos M.V.F., Lopes S.A., Lopes C.P., Machado J.A.,
 RA Machado M.A., Martins A.M.N., Madeira H.M.F., Marino S.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Meuck C.F.M., Miranda E.C., Miyaki R.Y., Monteiro-Vitorello C.R.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.F.O., Netto L.F.S.,
 RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E., Jr., de Sa P.G., Santelli P.V., Sawasaki H.E.,
 RA da Silva A.C.F., da Silva A.M., da Silva P., Silva W.A., Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Stuckira W.J., de Souza A.A.,
 RA de Souza A.P., Torontji M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis L., Schubel J.C.,
 RA "the genome sequence of the plant pathogen Xylella fastidiosa";
 RA Nature 406:151-159(2000).
 CC -1- SIMILARITY: TO HELICASE C-TERMINAL DOMAIN
 DE FMRL: AE003955; AAF84039.1;
 DE InterPro: IPR001410, DEAD;
 DE Pfam: PF00271, Helicase_C_1
 DE SMART: SM00490; HELIC_C_1
 KW ATP-binding, Complete proteome, Helicase.
 SQ SEQUENCE 833 AA, 94557 MW, 100AB0084C08AC8A Ck064,

Query Match 75.3%; Score 41; DB 16; Length 833;
 Best Local Similarity 75.3%; Pred. No. 54;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Q7 2 GLEHWPEL 9
 I:::I:::
 Db 656 GLEHWPEL 665

RESULT 6
 ID Q9P6Y1 PPPLIMINARY; PRT; 256 AA
 AC Q9P6Y1:
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN 13E11.350.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Dothideomycetes; Sordariomycetes.
 OC Sordariales; Sordariaceae; Neurospora
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte G., Ains W., Hobeisel J., Brandt P., Fartmann R., Holland P.,
 RA Nakamura G., Mewes H.W., Mannhaupt G.,
 RA Submitted (APR-2000) to the EMBL/GenBank/TrEMBL databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RL German Neurospora genome project,
 DE FMRL: AL353820; CAP88603.1; -
 DR InterPro: IPR001601; Meth-transf.
 KW Hypothetical protein.
 SQ SEQUENCE 256 AA, 27789 MW, 5251PR05856R000 Ck064.

Query Match 74.1%; Score 40; DB 3; Length 256;
 Best Local Similarity 76.3%; Pred. No. 24;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Q7 1 GLEHWPEL 9
 I:::I:::
 Db 141 GLEHWPEV 149

RESULT 7
 ID Q9KVI9 PPPLIMINARY; PRT; 285 AA.
 AC Q9KVI9:
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PHOSPHATIDYLSPHINGE DECARBOXYLASE.
 GN V00339.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 CX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX SUPA1N-EL 30K M5561; SKE01YPE 01;
 MEDLINE:20406833; PubMed:10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton K.A., Gwin M.L.,
 RA Dodson P.J., Haft D.H., Hickey P.K., Peterson T.P., Mayam I.A.,
 RA Gill S.E., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Emolaeva M.B., Vamathevan J., Bass S., Qin H., Bragol I., Sellers P.,
 RA McDonald L., Orntback T., Fleischmann P.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Gollwiler P., Melnick J.L., Venter A.C.,
 RA Fraser C.M.;
 RI "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 cholerae";
 RT Nature 406:477-483(2000).
 PL EMRL: AF004122; AAF93512.1; -
 DE TIGR: V00339; -
 DE InterPro: IPR000917; PS_harbaxylase.
 IF Pfam: PF02656; PS_harbaxylase;
 KW Complete proteome.
 SQ SEQUENCE 285 AA, 31651 MW, 5182A18FA173A38 Ck064,

Query Match 74.1%; Score 40; DB 16; Length 285;
 Best Local Similarity 62.5%; Pred. No. 27;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Q7 1 GLEHWPE 8
 I:::I:::
 Db 7 GLEWIPQ 14

RESULT 8
 ID O11317 PRELIMINARY; PRT; 337 AA.
 AC O11317:
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE H1-60.1 PROTEIN (FRAGMENT).
 GN H1-60.1
 OS Molluscum contagiosum virus subtype 1 (MCV1).
 OC Viruses; DNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Molluscipoxvirus.
 OX NCBI_TaxID=10280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Martilla M., Agreffer M., Nance A., Funes J.M., Varas A.J.,
 RA Lopez Esteban J.L., Esteban M., Martin-Gallardo A.,
 RA "A Random DNA Sequencing, Computer Based Approach for the Generation
 of a Gene Map of Molluscum Contagiosum Virus".
 PL Virus Genes 0:0-0(1997)
 DE FMRL: H86600; AAR57947.1; -
 DT NONTER 337 337
 SQ SEQUENCE 337 AA, 38567 MW, 1028090C2/866/RE Ck064;

```

Query Match          74.1%; Score 40; DB 12; Length 337;
Best Local Similarity 85.7%; Pred. No. 32;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 QHWPEL 9
Db 52 RHWPEL 58

RESULT 9
Q9X1Q4 PRELIMINARY: PRT: 452 AA.
AC Q9X1Q4
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE SIMILAR TO FLAVONOL 3-O-GLUCOSYLTRANSFERRASE.
GN FL011.22.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; eudicotyledons; Core eudicot; Rosidae.
OC eurosids 11; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=1702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN cv. COLUMBIA;
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altati H., Araujo P., Bolivar I., Bowley P., Gschler E., Dunn P.,
RA Gonzalez A., Kremenetskaia I., Kim C., Lee C., Li J., Li S.,
RA Luros S., Schwartz J., Shinn P., Tortum M., Vysotskaia V.S.,
RA Walker M., Yu C., Ecker J., Thellogis A., Davis R.W.
RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC006144; AAD04266.1;
KW Transferrase.
SQ SEQUENCE 452 AA; 50554 MW; 0E5E9C0211B6411C9C64.

Query Match          74.1%; Score 40; DB 10; Length 452;
Best Local Similarity 75.0%; Pred. No. 43;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LQHWPEL 9
Db 115 LAHWPEL 122

RESULT 10
Q33081 PRELIMINARY: PRT: 480 AA.
AC Q33081
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE HYPOHETICAL 51.4 KDA PROTEIN.
GN MLCB628.10C.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteriales; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Eigmeier K., Garnier T., De Rossi E., Psithi H., Cole S.T.;
RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93188700; PubMed=8446027;
RA Eigmeier K., Honore N., Woods S.A., Caudron R., Cole S.T.;
RT "Use of an ordered cosmid library to deduce the genomic organization
of Mycobacterium leprae."
RL Mol. Microbiol. 7:197-206(1993).
DR EMBL: V14967; CAA75198.1;
KW Hypothetical protein.
SQ SEQUENCE 480 AA; 51400 MW; 34B19B4M5674F55F C9C64;

Query Match          74.1%; Score 40; DB 16; Length 512;
Best Local Similarity 85.7%; Pred. No. 48;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GLOHWVP 7
Db 247 GYQHWVP 253

RESULT 11
Q9CDD9 PRELIMINARY: PRT: 512 AA.
AC Q9CDD9
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE PUTATIVE MEMBRANE PROTEIN.
GN MLC0047.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteriales; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN TN;
EX MEDLINE 21128732; PubMed 11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.P., Honore N., Garnier T., Churcher C., Harris D.,
RA Maudall K., Basham D., Brown D., Chillingworth T., Connor K.,
RA Davies K.M., Devlin K., Doherty S., Feltham T., Fraser A., Hamlin N.,
RA Hailloy S., Horneby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Sauter L., Seeger K., Simon S., Simmonds M., Skellon J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Bartrell B.G.;
EX "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
EX EMBL: AC583527; CAC25555.1;
DR Leprosoma; MLC047;
KW Complete proteome.
SQ SEQUENCE 512 AA; 54503 MW; FF7435DB39D6A800 C9C64;

Query Match          74.1%; Score 40; DB 16; Length 512;
Best Local Similarity 85.7%; Pred. No. 48;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GLOHWVP 7
Db 279 GYQHWVP 285

RESULT 12
Q98200 PRELIMINARY: PRT: 748 AA.
AC Q98200
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE MLC012L.
GN MLC012L.
OS Mollusca; Contagiosum virus subtype 1 (MCV1);
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Molluscipoxvirus.
OX NCBI_TaxID=10280;
RN [1]
RP SEQUENCE FROM N.A.
EX MEDLINE 96225459; PubMed 8670425;
RA Seckewitz T.G., Bugert J.J., Sisler J.R., Koonin E.V., Dairai G.,
RA Moss B.;
RT "Genome sequence of a human tumorigenic poxvirus: Prediction of

```


RT Specific host response-evasion genes";
PL Science 273:833-836(1996).

RN [2]
RP SEQUENCE FROM N.A.
RA Senkevitch T.G., Rupert J.J., Sisler J.P., Koonin E.V., Barai G.,
RA Moss B.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U60315, AAT55160.1,
SQ SEQUENCE 742 AA: 94411 MW: 61A27411F62CC6A5BA CRC64;

Query Match 74.1%; Score 40; DB 12; Length 748;
Best Local Similarity 62.5%; Pred. No. 71;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LQHWVPEL 9
Db 52 LQHWVPEL 59

RESULT 13
Q93708 PRELIMINARY: PRT: 1005 AA.
AC Q93708;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DI 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE TRANSPOSASE.
OS Escherichia coli.
OC Plasmid pMS001
OC Bacteria: Proteobacteria, gamma subdivision, Enterobacteriaceae
OC Escherichia.
OX NCBI_TaxID 562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=785-D; TRANSPOSON-IN60;
RX MEDLINE=20416819; PubMed=10858363;
RA Sabate M., Tarrago R., Navarro F., Miro E., Verges G., Barbe J.,
RA Prats G.;
RT "Cloning and sequence of the gene encoding a novel *ColE1*-like-
hybridizing bacteriophage phiX M-92 from *Escherichia coli* strain
RL Antimicrob. Agents Chemother. 44:1970-1973(2000)
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=785-D; TRANSPOSON-IN60;
RA Sabate M., Navarro F., Barbe J., Miro E., Mirelis B., Prats G.;
RT "A novel complex *col* type integron (In66) in *Escherichia coli*
PI carrying the *bla*CTX-M-9";
FI Submitted (oct-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF174129; AAK60193.1;
KW Plasmid.
SQ SEQUENCE 1005 AA: 116157 MW: 726E107EFAD057E CRC64;

Query Match 74.1%; Score 40; DB 2; Length 1005;
Best Local Similarity 62.5%; Pred. No. 96;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LQHWVPEL 9
Db 234 LQHWVPEL 241

RESULT 14
Q947B5 PRELIMINARY: PRT: 61 AA.
AC Q947B5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DI 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CRYPTOCROME 2 (FRAGMENT).
GN CRV2.
OS Cucumis melo (Muskmelon).

OC Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Fractheophyta;
OC Spermatophyta; Magnoliophyta, eudicotyledons; core eudicots; Rosidae;
OC eudicots I; Rubiales; Cucurbitaceae; Cucumis.
OX NCBI_TaxID 4656;
RN [1]
RP SEQUENCE FROM N.A.
RA Tarrago R., Sabate M., Relatase F., Barbe J., Guillamo G.;
RT "Tomato and barley contain duplicated copies of cryptochromes 1";
EL Plant Cell Environ. 24:991-997(2001).
DR EMBL: AF348457; AAL02088.1;
FT NON_TER 1
FT NON_TER 61
SQ SEQUENCE 61 AA: 7012 MW: 143262369059b CRC64;

Query Match 72.2%; Score 39; DB 10; Length 61;

Best Local Similarity 62.5%; Pred. No. 83;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LQHWVPEL 9
Db 48 LQHWVPEL 55

RESULT 15
Q9NM06 PRELIMINARY: PRT: 66 AA.
AC Q9NM06;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DI 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DI 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE PROBABLE RHO10.1 PROTEIN (FRAGMENT).
GN IM26.176.
OS Leishmania major.
OC Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID 5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FRIEDLIN;
RA Murphy J., Quail M., Harris B., Bajandream M., Lyons A., Barrell B.;
RT "Cloning and sequencing of the RHO10.1 gene from the LMP-1008 *Leishmania*
DR EMBL: AL160493; CAB97794.1;
FT NON_TER 66
SQ SEQUENCE 66 AA: 7496 MW: 7F66C0FA352A07C CRC64;

Query Match 73.2%; Score 39; DB 5; Length 66;
Best Local Similarity 62.5%; Pred. No. 9;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LQHWVPEL 9
Db 26 LQHWVPEL 33

Search completed: September 5, 2002, 15:31:06
Job time: 471 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:27:35 ; Search time: 58.86 seconds
(without alignments)
3.735 Million cell updates/sec

Title: US 09-744-804 39
Perfect score: 54
Sequence: 1 GLAHWPPEL 9

Scoring table: RUSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA*

- 1: /cgn2_6/ptodata/2/1aa/5A-6CMP.pep.*
- 2: /cgn2_6/ptodata/2/1aa/5B-6CMP.pep.*
- 3: /cgn2_6/ptodata/2/1aa/6A-6CMP.pep.*
- 4: /cgn2_6/ptodata/2/1aa/6B-6CMP.pep.*
- 5: /cgn2_6/ptodata/2/1aa/6C-6CMP.pep.*
- 6: /cgn2_6/ptodata/2/1aa/6D-6CMP.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	100.0	160	2	US-08-162-402B-10
2	54	100.0	160	2	US-08-162-402B-6
3	54	100.0	160	2	US-08-162-402B-8
4	34	72.2	160	2	US-08-162-402B-11
5	34	72.2	160	2	US-08-162-402B-20
6	34	72.2	160	2	US-08-162-402B-26
7	34	72.2	160	2	US-08-162-402B-9
8	34	72.2	160	2	US-08-162-402B-8
9	37	68.5	160	4	US-08-712-697-8
10	37	68.5	160	4	US-08-712-697-6
11	37	68.5	160	4	US-08-712-697-6
12	37	68.5	160	4	US-08-712-697-6
13	36	66.7	160	3	PCT-US92-03222-17
14	36	66.7	160	3	US-08-481-985B-110
15	36	66.7	160	3	US-08-481-985B-110
16	36	66.7	160	3	US-08-481-985B-113
17	36	66.7	160	3	US-08-481-985B-113
18	36	66.7	160	3	US-08-481-985B-113
19	36	66.7	160	4	US-08-960-190A-25
20	36	66.7	160	1	US-07-872-644-17
21	36	66.7	160	1	US-08-297-494-17
22	36	66.7	160	1	US-08-297-494-17
23	36	66.7	160	1	US-08-479-532-17
24	36	66.7	160	1	US-08-455-525-17
25	36	66.7	160	1	US-08-455-525-17
26	36	66.7	160	1	US-08-130-461-17
27	36	66.7	160	5	PCT-US92-03222-17

28 36 66.7 530 1 US-07-872-644-6
29 36 66.7 530 1 US-08-297-494-6
30 36 66.7 530 1 US-08-297-494-6
31 36 66.7 530 1 US-08-479-532-6
32 36 66.7 530 1 US-08-455-525-6
33 36 66.7 530 1 US-08-455-525-6
34 36 66.7 530 1 US-08-130-461-6
35 36 66.7 530 1 US-08-130-461-6
36 36 66.7 530 1 US-07-872-644-49
37 36 66.7 530 1 US-08-297-494-49
38 36 66.7 530 1 US-08-297-494-49
39 36 66.7 530 1 US-08-479-532-49
40 36 66.7 530 1 US-08-455-525-49
41 36 66.7 530 1 US-08-455-525-49
42 36 66.7 530 1 US-08-130-461-49
43 36 66.7 530 5 PCT-US92-03222-17
44 36 66.7 674 1 US-08-130-461-17
45 36 66.7 674 1 US-08-803-972-7

ALIGNMENTS

RESULT 1
US-08-162-402B-10
Sequence 10, Application US/98162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERIAI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LAIOCCA, DAVID J.
TITLE OF INVENTION: 46 KUALIT HUMAN MILK FAT
TITLE OF INVENTION: GLOBULE (HMEG) ANIGEN, FRAGMENT & FUSION PROTEIN
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESS: Priddy, Schroeder & Poplawski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER FEATURES FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/98162402B
FILING DATE: 03 Sep 1993
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
FILING DATE:
ATTORNEY/AGENCY INFORMATION:
NAME: Amzel, Viviana
REFERENCE NUMBER: 30,939
REFERENCE/BACKUP NUMBER: PCT 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 313-469-4310
TELEFAX: 313-469-4310
INFORMATION FOR SEQ ID NO. 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 160 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-10

Query Match: 100.0% Score 54; 160 Length 160
Best Local Similarity 100.0% Pred. No. 0.015;

```

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLOHWVPEL 9
   | | | | | | | |
Db 29 GLOHWVPEL 37

RESULT 2
US-08-162-402B-6
; Sequence 6, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANT, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: IAROCCA, DAVID J.
; TITLE OF INVENTION: 46 KDAUTON HUMAN MILK FAT
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSPQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 387 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-162-402B-6

Query Match 100.0%; Score 54; DB 2; Length 387;
Best Local Similarity 100.0%; Pred No. 0.034;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLOHWVPEL 9
   | | | | | | | |
Db 97 GLOHWVPEL 105

RESULT 3
US-08-162-402B-8
; Sequence 8, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANT, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: IAROCCA, DAVID J.
; TITLE OF INVENTION: 46 KDAUTON HUMAN MILK FAT
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSPQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 387 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-162-402B-8

```

```

; TITLE OF INVENTION: GLOBULE (HMF3) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSPQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 465 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-162-402B-8

Query Match 100.0%; Score 54; DB 2; Length 465;
Best Local Similarity 100.0%; Pred No. 0.047;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GLOHWVPEL 9
   | | | | | | | |
Db 175 GLOHWVPEL 183

RESULT 4
US-08-162-402B-11
; Sequence 11, Application US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANT, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: IAROCCA, DAVID J.
; TITLE OF INVENTION: 46 KDAUTON HUMAN MILK FAT
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSPQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B

```

1 FILING DATE: 03-DEC-1993
2 CLASSIFICATION: 435
3 PRIOR APPLICATION DATA:
4 APPLICATION NUMBER:
5 FILING DATE:
6 ATTORNEY/AGENT INFORMATION:
7 NAME: Anzel, Violaana
8 REGISTRATION NUMBER: 30,930
9 REFERENCE/DOCKET NUMBER: P66 38215
10 TELECOMMUNICATION INFORMATION:
11 TELEPHONE: 213-622-7700
12 TELEFAX: 213-489-4210
13 TELEX:
14 INFORMATION FOR SEQ ID NO: 11:
15 SEQUENCE CHARACTERISTICS:
16 LENGTH: 160 amino acids
17 TYPE: amino acid
18 STRANDEDNESS: unknown
19 TOPOLOGY: unknown
20 MOLECULE TYPE: peptide
21 US-08-162-402B-11

Query Match 72.2%; Score 39; DP 2; Length 160;
Best Local Similarity 77.8%; Pred. No. 6.7;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GLQHWVPEL 9
DB 29 GLQHWVPEL 37

RESULT 5
US-08-480-229C-20
1 Sequence 20, Application US/08480229C
2 Patent No. 5874562
3 GENERAL INFORMATION:
4 APPLICANT: Quettermous, Thomas
5 APPLICANT: Hojan, Brigid
6 APPLICANT: Snodgrass, H. Ralph
7 APPLICANT: Zupancic, Thomas J.
8 TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
9 TITLE OF INVENTION: CELL LOCUS 1
10 NUMBER OF SEQUENCES: 29
11 CORRESPONDENCE ADDRESS:
12 ADDRESSEE: Pennie & Edmonds LLP
13 STREET: 1155 Avenue of the Americas
14 CITY: New York
15 STATE: New York
16 COUNTRY: United States
17 ZIP: 10036-2711
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Floppy disk
20 COMPUTER: IBM PC compatible
21 OPERATING SYSTEM: PC DOS/MS-DOS
22 SOFTWARE: PatentIn Release #1.0, Version #1.30
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US 08/480,229C
25 FILING DATE: 07-JUN-1995
26 CLASSIFICATION: 536
27 ATTORNEY/AGENT INFORMATION:
28 NAME: Poissant, Brian M.
29 REGISTRATION NUMBER: 28,462
30 REFERENCE/DOCKET NUMBER: 8407-0026-999
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE: (212) 790-9040
33 TELEFAX: (212) 869-8864/9741
34 TELEX: 66141 Pennie
35 INFORMATION FOR SEQ ID NO: 20:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 320 amino acids
38 TYPE: amino acid
39 STRANDEDNESS:

1 TOPOLOGY: unknown
2 MOLECULE TYPE: protein
3 US 08 480-229C-20

Query Match 72.2%; Score 39; DP 2; Length 320;
Best Local Similarity 77.8%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GLQHWVPEL 9
DB 28 GLQHWVPEL 46

RESULT 6
US-08-659-235C-20
1 Sequence 20, Application US/08659235C
2 Patent No. 5977281
3 GENERAL INFORMATION:
4 APPLICANT: Quettermous, Thomas
5 APPLICANT: Hojan, Brigid
6 APPLICANT: Snodgrass, H. Ralph
7 APPLICANT: Zupancic, Thomas J.
8 TITLE OF INVENTION: DEVELOPMENTALLY REGULATED ENDOTHELIAL
9 TITLE OF INVENTION: CELL LOCUS-1
10 NUMBER OF SEQUENCES: 29
11 CORRESPONDENCE ADDRESS:
12 ADDRESSEE: Pennie & Edmonds LLP
13 STREET: 1155 Avenue of the Americas
14 CITY: New York
15 STATE: New York
16 COUNTRY: United States
17 ZIP: 10036-2711
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Floppy disk
20 COMPUTER: IBM PC compatible
21 OPERATING SYSTEM: PC DOS/MS-DOS
22 SOFTWARE: PatentIn Release #1.0, Version #1.30
23 CURRENT APPLICATION DATA:
24 APPLICANT: Quettermous, Thomas
25 FILING DATE: 05-JUN-1996
26 CLASSIFICATION: 435
27 ATTORNEY/AGENT INFORMATION:
28 NAME: Poissant, Brian M.
29 REGISTRATION NUMBER: 28,462
30 REFERENCE/DOCKET NUMBER: 8907-0034-999
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE: (212) 790-9040
33 TELEFAX: (212) 869-8864/9741
34 TELEX: 66141 Pennie
35 INFORMATION FOR SEQ ID NO: 20:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 320 amino acids
38 TYPE: amino acid
39 STRANDEDNESS:
40 TOPOLOGY: unknown
41 MOLECULE TYPE: protein
42 US-08-659-235C-20

Query Match 72.2%; Score 39; DP 2; Length 320;
Best Local Similarity 77.8%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GLQHWVPEL 9
DB 28 GLQHWVPEL 46

RESULT 7
US-08-162-402B-9
1 Sequence 9, Application US/08162402B
2 Patent No. 5972347

```

; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: IARCCA, DAVID I.
; TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
; TITLE OF INVENTION: GLOBUF (HMG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pctty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/98/162,402B
; FILING DATE: 03-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,940
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7709
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 463 amino acids
; TYPE: amino acid
; STRAIN/GENESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-162-402B-9

Query Match 72.2%; Score 39; DB 2; Length 463;
Best Local Similarity 77.8%; Pred. No. 21;
Matches 7; Conservative 9; Mismatches 2; Indels 9; Gaps 0.

QY 1 GLOHWVPEL 9
DB 175 GLOHWGPEL 183

; GENERAL INFORMATION:
; APPLICANT: WHITNEY, Michael A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SENSITIVE
; TITLE OF INVENTION: AND RAPID, FUNCTIONAL IDENTIFICATION OF GENOMIC
; TITLE OF INVENTION: POLYNUCLEOTIDES AND SEQUENCE SCREENING CAPABILITIES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/719 697
; FILING DATE: 26-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Halls, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 08366/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 264 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-719-697-8

Query Match 68.5%; Score 47; DB 2; Length 264;
Best Local Similarity 75.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0.

QY 2 LOHWVPEL 9
DB 138 LOHWPEL 145

; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Zickarnik, Gregor
; TITLE OF INVENTION: SUBSTRATES FOR BETA-LACTAMASE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/727,616A
; FILING DATE: 15-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/467,544
; FILING DATE: 20-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Halls, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 08367/0034001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 264 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-727-616A-8

```

Query Match 68.5%; Score 37; DB 4; Length 264;
 Best Local Similarity 75.0%; Pred. No. 26;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 LQHWPEL 9
 1 11 1111

DB 139 LQHWPEL 145

RESULT 10
 US-08-719-697-6
 ; Sequence 6, Application US/08719697
 ; Patent No. 5928828
 ; GENERAL INFORMATION:
 ; APPLICANT: Whitley, Michael A.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SENSITIVE
 ; TITLE OF INVENTION: AND RAPID, FUNCTIONAL IDENTIFICATION OF GENOMIC
 ; TITLE OF INVENTION: POLYNUCLEOTIDES AND SEQUENTIAL SCREENING DATA
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Fish & Richardson P.C.
 ; STREET: 4225 Executive Square, Suite 1400
 ; CITY: La Jolla
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 92037
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/719,697
 ; FILING DATE: 26-SEP-1996
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hallie, Lisa A.
 ; REGISTRATION NUMBER: 48,347
 ; REFERENCE: OFFICE REFERENCE: 0003001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 619/678-5070
 ; TELEFAX: 619/678-5070
 ; INFORMATION FOR SEQ ID NO: 6:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 265 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-719-697-6

Query Match 68.5%; Score 37; DB 2; Length 265;
 Best Local Similarity 75.0%; Pred. No. 26;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 LQHWPEL 9
 1 11 1111

DB 139 LQHWPEL 146

RESULT 11
 US-08-727-616A-6
 ; Sequence 6, Application US/08727616A
 ; Patent No. 6291162
 ; GENERAL INFORMATION:
 ; APPLICANT: Tsien, Roger Y.
 ; APPLICANT: Zlokovic, George
 ; TITLE OF INVENTION: SUBSTRATES FOR BETA-LACTAMASE
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Fish & Richardson P.C.

STREET: 4225 Executive Square, Suite 1400
 CITY: La Jolla
 STATE: CA
 COUNTRY: USA
 ZIP: 92037
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows 95
 SOFTWARE: FASTSEQ for Windows Version 2.0b
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/727,616A
 FILING DATE: 15-OCT-1996
 PRIOR APPLICATION DATA: 08/407,544
 APPLICATION NUMBER: 20-MAR-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Hallie, Lisa A.
 REGISTRATION NUMBER: 48,347
 REFERENCE: OFFICE REFERENCE: 0003001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619/678-5070
 TELEFAX: 619/678-5070
 INFORMATION FOR SEQ ID NO: 6:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 265 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 US-08-727-616A-6

Query Match 68.5%; Score 37; DB 4; Length 265;
 Best Local Similarity 75.0%; Pred. No. 26;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 LQHWPEL 9
 1 11 1111

DB 139 LQHWPEL 146

RESULT 12
 PCT-US94-00198-4
 ; Sequence 4, Application PC/TUS9400198
 ; GENERAL INFORMATION:
 ; APPLICANT: Schering Corp.
 ; TITLE OF INVENTION: RAS Associated GAP Proteins
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Schering Corp.
 ; STREET: 1 Giraldo Farms
 ; CITY: Madison
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 04304-1104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: Macintosh
 ; OPERATING SYSTEM: 6.0.8
 ; SOFTWARE: Microsoft Word 5.1a
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US94/00198
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/004,824
 ; FILING DATE: 15-JAN-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Iann, Paul G.
 ; REGISTRATION NUMBER: 42,743
 ; REFERENCE: OFFICE REFERENCE: 000352 PCT
 ; TELECOMMUNICATION INFORMATION:

```

: TELEPHONE: (201)822-7255
: TELEFAX: (201)822-7039
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3079 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Saccharomyces cerevisiae
: PCT-US94-00198-4

Query Match 68.5%; Score 37, DB 5, Length 3079,
Best Local Similarity 65.6%; Pred. No. 3;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 LQHWPEL 9
I:|:|:|
Db 2418 CLSYWIPNL 2426

RESULT 14
US-08-484-905-110
: Sequence 110, Application US/08484905
: Patent No. 5976551
: GENERAL INFORMATION:
: APPLICANT: Mottez, Estelle
: APPLICANT: Abastado, Jean-Pierre
: TITLE OF INVENTION: An Altered Major Histocompatibility
: TITLE OF INVENTION: Complex(MHC) Determinant and Methods for Using the
: TITLE OF INVENTION: Determinant
: NUMBER OF SEQUENCES: 127
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
: STREET: 1300 I Street, N.W., Suite 700
: CITY: Washington
: STATE: D.C.
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/484,905
: FILING DATE: 07-JUNE-1995
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/801,818
: FILING DATE: 05-DEC-1991
: CLASSIFICATION: 530
: INFORMATION FOR SEQ ID NO: 110:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 255 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-484-905-110

```

```

Query Match 66.7%; Score 36, DB 2, Length 255;
Best Local Similarity 62.5%; Pred. No. 38;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LQHWPEL 9
I:|:|:|
Db 201 LKHWPEI 208

RESULT 14
US-08-481-985B-110
: Sequence 110, Application US/08481985B
: Patent No. 6011146
: GENERAL INFORMATION:
: APPLICANT: Mottez, Estelle
: APPLICANT: Abastado, Jean-Pierre
: APPLICANT: Kourilsky, Philippe
: TITLE OF INVENTION: Altered Major Histocompatibility Complex
: TITLE OF INVENTION:
: NUMBER OF SEQUENCES: 148
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
: STREET: 1300 I Street, N.W., Suite 700
: CITY: Washington
: STATE: D.C.
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/481,985B
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/801,818
: FILING DATE: 05-DEC-1991
: CLASSIFICATION: 435
: INFORMATION FOR SEQ ID NO: 110:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 255 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-481-985B-110

```

```

Query Match 66.7%; Score 36, DB 3, Length 255;
Best Local Similarity 62.5%; Pred. No. 38;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LQHWPEL 9
I:|:|:|
Db 201 LKHWPEI 208

RESULT 15
US-08-370-476-110

```


; Sequence 110, Application US/08370476
; Patent No. 6153488
; GENERAL INFORMATION:
; APPLICANT: Mottez, Estelle
; APPLICANT: Abastado, Jean-Pierre
; APPLICANT: Kourilsky, Philippe
; APPLICANT: Lone, Yu-Chun
; APPLICANT: Ojcius, David
; APPLICANT: Carrouge, Amanda
; TITLE OF INVENTION: Altered Major Histocompatibility Complex
; NUMBER OF SEQUENCES: 127
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSER: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,476
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/117,575
; FILING DATE: 07-SEP-1993
; APPLICATION NUMBER: US 08/072,787
; FILING DATE: 06-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/801,818
; FILING DATE: 05-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/792,473
; FILING DATE: 15-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/LAW FIRM: 05243 0001-01000
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4490
; INFORMATION FOR SEQ ID NO: 110:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-370-476-110

Query Match 66.7%; Score 36; DB 4; Length 255;
Best Local Similarity 62.5%; Pred. No. 38;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LOHWPEL 9
|:|:|:|:
Db 201 LKHWPEI 208

Search completed: September 5, 2002, 15:27:36
Job time: 321 sec

QY 1 VOFVASYKV 9
 Db 24 VOFVASYKV 32

RESULT 2

US-08-659-235C-2

Sequence 2, Application US/08/59235C
 Patent No. 5877281
 GENERAL INFORMATION:
 APPLICANT: Quertermous, Thomas
 APPLICANT: Hogan, Bridg
 APPLICANT: Shodgrass, H. Ralph
 APPLICANT: Zupancic, Thomas J.
 TITLE OF INVENTION: DEVELOPMENTALLY REGULATED ENDOTHELIAL
 TITLE OF INVENTION: CELL LOCUS 1
 NUMBER OF SEQUENCES: 29
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds LLP
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: United States
 ZIP: 10046-2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: 03/28/55,235;
 FILING DATE: 05-JUN-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Poissant, Brian M.
 REGISTRATION NUMBER: 28,452
 REFERENCE/DOCKET NUMBER: 8607-0044-604
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741
 TELEX: 66141 penio
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 85 amino acids
 TYPE: amino acid
 STRANDEDNESS: unknown
 TOPOLOGY: peptide
 MOLECULE TYPE: peptide
 US-08-659-235C-2

Query Match 100.0%; Score 43; DB 2; Length 85;
 Best Local Similarity 100.0%; Prod. No. 0.03;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VOFVASYKV 9
 Db 24 VOFVASYKV 32

RESULT 3

US-08-162-402B-12

Sequence 12, Application US/08/162402B
 Patent No. 5972347
 GENERAL INFORMATION:
 APPLICANT: CERIANI, ROBERTO L.
 APPLICANT: PETERSON, JERRY A.
 APPLICANT: LAROCCA, DAVID J.
 TITLE OF INVENTION: 46 KDa HUMAN MILK FAT
 TITLE OF INVENTION: GLOBULE (HMG) ANTIGEN, FRAGMENTS & PEPTIDE
 NUMBER OF SEQUENCES: 29
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pretty, Schroeder & Poplawski

STREET: 444 South Flower St., 19th Floor
 CITY: Los Angeles
 STATE: CA
 COUNTRY: USA
 ZIP: 90071
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/162,402B
 FILING DATE: 03-DEC-1993
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Amzel, Viviana
 REGISTRATION NUMBER: 30,930
 REFERENCE/DOCKET NUMBER: P66 38215
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 213-622-7700
 TELEFAX: 213-489-4210
 TELEX:
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 159 amino acids
 TYPE: amino acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 US-08-162-402B-12

Query Match 100.0%; Score 43; DB 2; Length 159;
 Best Local Similarity 100.0%; Prod. No. 0.058;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VOFVASYKV 9
 Db 85 VOFVASYKV 93

RESULT 4

US-07-607-538C-3

Sequence 3, Application US/07/607538C
 Patent No. 5455031
 GENERAL INFORMATION:
 APPLICANT: Ceriani Dr., Roberto L.
 APPLICANT: Peterson Dr., Jerry A.
 APPLICANT: Larocca, David J.
 TITLE OF INVENTION: POLYPEPTIDE WITH 46
 TITLE OF INVENTION: DIFFERENTIATION ANTIGEN BINDING SPECIFICITY AND
 TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
 TITLE OF INVENTION: FUSION PROTEIN, POLYNUCLEOTIDE AND POLYPRO-
 TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-
 TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF
 TITLE OF INVENTION: USE THEREOF
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: V. Amzel & Assoc.
 STREET: 2355 No. 5455031th Broadway
 CITY: Walnut Creek
 STATE: California
 COUNTRY: USA
 ZIP: 94596
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS 5.0
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US-07-607,538C
FILING DATE: 01-NOV-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Viviana Amzel
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: CHFC-004
TELEPHONE: (510) 943-1931
TELEFAX: (510) 943-1189
TELEX: N.A.
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 217 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE:
US-07-607-538C-3

Query Match 100.0%; Score 43; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 0.081;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VQFVASYKV 9
DB 143 VQFVASYKV 151

RESULT 5
US-08-162-402R-3
Sequence 3, Application US/08162402R
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERIANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LARocca, DAVID J.
TITLE OF INVENTION: 46 KILTON HUMAN MILK FAT
TITLE OF INVENTION: GLOBULE (HMFQ) ANTIGEN, FRAGMENT & FUSION PROTEIN
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplawski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US-08-162,402R
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: P06 38215
TELEPHONE: 213-622-7780
TELEFAX: 213-489-4210
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 217 amino acids
TYPE: amino acid

STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402R-3

Query Match 100.0%; Score 43; DB 2; Length 217;
Best Local Similarity 100.0%; Pred. No. 0.081;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VQFVASYKV 9
DB 143 VQFVASYKV 151

RESULT 6
US-07-607-538C-2
Sequence 2, Application US/07607538C
Patent No. 5455031
GENERAL INFORMATION:
APPLICANT: Ceriani Dr., Roberto L.
APPLICANT: Peterson Dr., Jerry A.
APPLICANT: Larocca, David J.
TITLE OF INVENTION: DIFFERENTIATION WITH 46
TITLE OF INVENTION: FACTORS V AND VIII LIGHT-CHAIN HOMOLOGIES,
TITLE OF INVENTION: FUSION PROTEIN, POLYNUCLEOTIDE AND POLYPEPTIDE
TITLE OF INVENTION: NUCLEOTIDE ENCODING THE POLYPEPTIDE, ANTI-
TITLE OF INVENTION: POLYPEPTIDE ANTIBODIES, KITS AND METHODS OF
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: V. Amzel & Assoc.
STREET: 2055 No. 5455031th Broadway
CITY: Walnut Creek
STATE: California
COUNTRY: USA
ZIP: 94596
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 5.0
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07-607,538C
FILING DATE: 01-NOV-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Viviana Amzel
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: CHFC-004
TELEPHONE: (510) 943-1931
TELEFAX: (510) 943-1189
TELEX: N.A.
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE:
US-07-607-538C-2

Query Match 100.0%; Score 43; DB 1; Length 218;
Best Local Similarity 100.0%; Pred. No. 0.081;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VQFVASYKV 9
DB 144 VQFVASYKV 152

```

RESULT 7
US-08-162-402B-2
; Sequence 7, Application: US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LARocca, DAVID J.
; TITLE OF INVENTION: 46 KLALTON HUMAN MILK FAT
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-162-402B-2

Query Match 100.0%; Score 43; PR 2; Length 218;
Best Local Similarity 100.0%; Prod No. 0.081;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VOFVASYKV 9
Db 144 VOFVASYKV 152

RESULT 8
US-08-162-402B-6
; Sequence 8, Application: US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LARocca, DAVID J.
; TITLE OF INVENTION: 46 KLALTON HUMAN MILK FAT
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor

```

```

; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amzel, Viviana
; REGISTRATION NUMBER: 30,930
; REFERENCE/DOCKET NUMBER: P66 38215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 387 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-162-402B-6

Query Match 100.0%; Score 43; PR 2; Length 387;
Best Local Similarity 100.0%; Prod No. 0.15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VOFVASYKV 9
Db 313 VOFVASYKV 321

RESULT 9
US-08-162-402B-8
; Sequence 9, Application: US/08162402B
; Patent No. 5972337
; GENERAL INFORMATION:
; APPLICANT: CERIANI, ROBERTO L.
; APPLICANT: PETERSON, JERRY A.
; APPLICANT: LARocca, DAVID J.
; TITLE OF INVENTION: 46 KLALTON HUMAN MILK FAT
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder & Poplawski
; STREET: 444 South Flower St., 19th Floor
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/162,402B
; FILING DATE: 03-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:

```

ATTORNEY/AGENT INFORMATION:
NAME: Amzel, Viorana
REGISTRATION NUMBER: 40,940
REFERENCE/DOCKET NUMBER: P66 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 465 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-8

Query Match 100.0%; Score 43; DB 2; Length 465;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VQFVASYKV 9
DB 391 VQFVASYKV 399

RESULT 10
US-08-480-229C-21
Sequence 21, Application US/09490229C
Patent No. 5874562
GENERAL INFORMATION:
APPLICANT: Quentormous, Thomas
APPLICANT: Hogan, Brigid
APPLICANT: Snodgrass, H. Ralph
TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
TITLE OF INVENTION: CELL LOCUS-1
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08-480-229C
APPLICATION NUMBER: 08-480-229C
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0026-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 Pennie
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 321 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: protein
MOLECULE TYPE: protein
US-08-480-229C-21

US-08-480-229C-21
Sequence 21, Application US/09490229C
Patent No. 5874562
GENERAL INFORMATION:
APPLICANT: Quentormous, Thomas
APPLICANT: Hogan, Brigid
APPLICANT: Snodgrass, H. Ralph
TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
TITLE OF INVENTION: CELL LOCUS-1
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08-480-229C
APPLICATION NUMBER: 08-480-229C
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0026-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 Pennie
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 321 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-480-229C-21

Query Match 83.7%; Score 36; DB 2; Length 321;
Best Local Similarity 77.8%; Pred. No. 4.6;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VQFVASYKV 9
DB 246 VQFVASYKV 254

RESULT 11
US-08-659-235C-21
Sequence 21, Application US/08659235C
Patent No. 5877281
GENERAL INFORMATION:
APPLICANT: Quentormous, Thomas
APPLICANT: Hogan, Brigid
APPLICANT: Snodgrass, H. Ralph
TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
TITLE OF INVENTION: CELL LOCUS-1
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08-659-235C
APPLICATION NUMBER: US/08-659-235C
FILING DATE: 05-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0034-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 Pennie
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 321 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-659-235C-21

Query Match 83.7%; Score 36; DB 2; Length 321;
Best Local Similarity 77.8%; Pred. No. 4.6;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VQFVASYKV 9
DB 246 VQFVASYKV 254

RESULT 12
US-08-480-229C-10
Sequence 10, Application US/08490229C
Patent No. 5874562
GENERAL INFORMATION:
APPLICANT: Quentormous, Thomas
APPLICANT: Hogan, Brigid
APPLICANT: Snodgrass, H. Ralph
TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
TITLE OF INVENTION: CELL LOCUS-1
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08-480-229C
APPLICATION NUMBER: 08-480-229C
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0026-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 Pennie
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 321 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-480-229C-21

```

; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; TITLE OF INVENTION: CELL LOCUS-1
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036 2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,229C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/BOOKET NUMBER: 8907-0026-999
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-229C-10

```

```

Query Match      83.7%  Score 36, DB 2, Length 480;
Best Local Similarity 77.8%  Pred. No. 5 5;
Matches 7; Conservative 1, Mismatches 1, Indels 0, Gaps 0;

```

```

QY 1 VQFVASYKV 9
    |||||
DB 402 VQFVGSYKL 410

```

```

RESULT 13
US-08-659-235C-10
; Sequence 10, Application US/08/659235C
; Patent No. 5877281
; GENERAL INFORMATION:
; APPLICANT: Quentermous, Thomas
; APPLICANT: Hogan, Briqid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036 2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,235C
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 435

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/BOOKET NUMBER: 8907-0026-999
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-659-235C-10

```

```

Query Match      83.7%  Score 36; DB 2; Length 480;
Best Local Similarity 77.8%  Pred. No. 5 5;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 VQFVASYKV 9
    |||||
DB 402 VQFVGSYKL 410

```

```

RESULT 14
US-08-480-229C-14
; Sequence 14, Application US/08/480229C
; Patent No. 5874562
; GENERAL INFORMATION:
; APPLICANT: Quentermous, Thomas
; APPLICANT: Hogan, Briqid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.40
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,229C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/BOOKET NUMBER: 8907-0026-999
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 Pennie
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 513 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-480-229C-14

```

```

Query Match      83.7%  Score 36; DB 2; Length 513;

```


Best local Similarity 77.8%; Pred. NO. 5.9;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VQFVASYKV 9
|||||
DB 435 VQFVGSYKL 443

RESULT 15
US-08-659-235C-14
: Sequence 14, Application US/08659235C
: Patent No. 5877281
: GENERAL INFORMATION:
: APPLICANT: Quertormous, Thomas
: APPLICANT: Hogan, Bridgid
: APPLICANT: Shodgrass, H. Ralph
: APPLICANT: Zupancic, Thomas J.
: TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
: TITLE OF INVENTION: CELL LOCUS-1
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pennie & Edmonds LLP
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: United States
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08-659-235C
: FILING DATE: 05-JUN-1996
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Poissant, Brian M.
: REGISTRATION NUMBER: 28,462
: REFERENCE/DOCKET NUMBER: 8907-0044-999
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864/9741
: TELEX: 66141 Pennie
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 513 amino acids
: TYPE: amino acid
: STRANDEDNESS: unknown
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-659-235C-14

Query Match 83.7%; Score 36; DB 2; Length 513;
Best local Similarity 77.8%; Pred. NO. 5.9;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VQFVASYKV 9
|||||
DB 435 VQFVGSYKL 443

Search completed: September 5, 2002, 15:27:36
Job time: 321 sec

XX WPI: 2000-205464/18.
 XX
 XX Tumor associated antigen peptides, especially derived from uroplakin,
 PT useful as vaccines to prevent or cure cancers including breast,
 PT bladder, prostate, pancreas, ovary, thyroid, colon and stomach
 XX
 XX Claim 17; Page 100; 113pp; English.
 XX
 XX Tumor associated antigen peptides (TAA) may be used for the
 CC treatment, prevention and cure of cancer or cancer metastases. The
 CC cancer may be breast, bladder, prostate, pancreas, ovary, thyroid,
 CC colon, stomach, head or neck cancer or a carcinoma. The tumor
 CC associated antigens are presentable to the immune system by HLA A2
 CC molecules and are generally between 8 to 10 amino acids in length.
 CC The amino acids located at positions 2 and 9 of the tumor associated
 CC antigens are the anchor residues which participate in the binding to
 CC MHC class I molecules, more specifically HLA-A2. More tumor
 CC associated antigens are described in GENESeq records: AAY82805-Y82882.
 CC Those tumor associated antigens described in records AAY82805-Y82882,
 CC and AAY82855-Y82869 are derived from uroplakin, such as Uroplakin II,
 CC Uroplakin Ia, Uroplakin II and Uroplakin Ib. These described in
 CC records AAY82825-Y82829 are derived from prostate specific antigen
 CC (PSA). Those described in records AAY82830-Y82835 are derived from
 CC prostate specific membrane antigen (PSMA). Those described in
 CC records Y82836 AAY82839 are derived from prostate acid phosphatase
 CC (PAP). Those described in records AAY82840-Y82846 are derived
 CC from lactadherin (LA-46). Those described in records AAY82847-Y82854
 CC are derived from Macin and those described in records AAY82871-Y82882
 CC are derived from Teratocarcinoma derived growth factor (GDF10-1).
 XX
 XX Sequence 9 AA:
 SQ

Query Match: 100.0%; Score 43; DB 21; Length 9;
 Best Local Similarity: 100.0%; Pred. No. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VQFVASYKV 9
 IIIIIIII
 Db 1 vqfvasykv 9

RESULT 2
 AA044690
 LU AAW94690 standard; peptide: 95 AA.
 XX
 XX AAW94690;
 AC

DT 05-MAY-1999 (first entry)
 XX
 XX Discoidin I/factor VIII like domain #2.
 DE
 XX
 FW Irel-1; developmentally regulated endothelial cell locus 1; cancer;
 KW discoidin I; factor VIII like domain; epidermal growth factor; EGF;
 KW diabetic retinopathy; rheumatoid arthritis; endometriosis;
 KW angiogenesis.
 XX
 XX Homo sapiens.
 OS
 OS Synthetic.

XX US5877281-A.
 XX
 XX 02-MAR-1999.
 XX
 XX 05-JUN-1996; 96US-0659235.
 XX
 XX 05-JUN-1996; 96US-0659235.
 XX
 XX 07-JUN-1996; 96US-0480229.
 XX
 XX (PROG-) PROGENITOR INC.
 PA (UYVA-) UNIV VANDERBILT.

XX

PI Hoan B, Quertermous T, Snodgrass BR, Zupancic TJ;
 XX WPI: 1999-189720/16.
 XX
 XX Del 1 polypeptide sequences useful for treatment of cancer,
 PT diabetic retinopathy, rheumatoid arthritis and endometriosis
 XX
 XX Example; Column 39-40; 73pp; English.
 XX
 XX The present sequence represents a discoidin I/factor VIII like domain.
 CC The present invention describes developmentally-regulated endothelial
 CC cell locus 1 (del 1). Del 1 has epidermal growth factor like domains and
 CC discoidin I/factor VIII-like domains. The del 1 proteins have an
 CC inhibitory effect on angiogenesis (blood vessel growth). This activity
 CC may be useful clinically to prevent neovascularization of tissues such
 CC as tumor nodules and prevention of metastases. The anti-angiogenic
 CC activity of del 1 may be used to treat abnormal conditions that result
 CC from angiogenesis, including cancer, diabetic retinopathy, rheumatoid
 CC arthritis and endometriosis. Since Del-1 promotes angiogenesis it can be
 CC used to treat ocular ischemia, choroidal neovascularization, wound healing and
 CC peripheral vascular disease. Del 1 is also useful for promoting bone
 CC formation. Del-1 binds to alpha V beta 3 integrin, and is an apoptosis
 CC inducer.
 XX
 XX Sequence 85 AA:
 SQ

Query Match: 100.0%; Score 43; DB 20; Length 85;
 Best Local Similarity: 100.0%; Pred. No. 0.13;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VQFVASYKV 9
 IIIIIIII
 Db 24 vqfvasykv 32

RESULT 3
 AA077253
 LU AAR77253 standard; Protein: 217 AA.
 XX
 XX AAR77253;
 AC

DT 21-NOV-1995 (first entry)
 XX
 XX HMEG 46 kDa antigen C terminal region.
 DE

XX HMEG; human milk fat globule; antigen; immunogen; vaccine; virucide;
 KW epithelium; tumor; breast cancer; monoclonal antibody; MAH.
 KW
 XX Homo sapiens.

XX W09515171-A.
 DN

XX 08-JUN-1995.
 PP

XX 05-DEC-1994; 94W0-0813967.
 PP

XX 03-DEC-1993; 93US-0162402.
 PP

XX (CANC-) CANCER RES FUND CONTRA COSTA.
 PA

XX Coriani RL, Larocca DJ, Peterson JA;
 PI

XX WPI: 1995-215151/28.
 XX

XX 46 kD apparent molecular weight human milk fat globule antigen -
 PT used in assays to determine the presence of a cancerous tumour of
 PT epithelial origin, and in a vaccine against neoplastic tumours
 XX
 XX Claim 7; Page 41; 68pp; English.

XX
 XX A partial cDNA clone BA46-1 (AA091199) for the 46 kDa HMEG antigen,
 CC a major component of the apical surface of the normal breast

CC epithelial cell, was obtained by PCR of cDNA from a lactating
 CC breast cDNA library. The C-terminal region of the encoded protein
 CC showed 43% identity to corresp. regions of human Factor-V and 38%
 CC to Factor-VIII.
 XX
 SQ Sequence 217 AA;

Query Match 100.0%; Score 43; DB 16; Length 217;
 Best Local Similarity 100.0%; Pred No. 0.33;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VQFVASYKV 9
 DB 143 vqfvasykv 151
 ||| |||||

RESULT 4
 AAR77254
 ID AAR77254 standard; Protein; 218 AA.
 XX
 AC AAR77254;
 XX
 DT 21-NOV-1995 (first entry)
 XX
 DE HMFG 46 kDa antigen partial sequence.
 XX
 KW HMFG; human milk fat globule; antigen; immunogen; vaccine; virucide;
 KW epithelium; tumor; breast cancer; monoclonal antibody; MAb.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Modified-site 69..71
 FT /label= N-glycosylation_site
 FT Modified-site 156..158
 FT /label= N-glycosylation_site
 FT Modified-site 160..162
 FT /label= N-glycosylation_site
 FT Modified-site 181..183
 FT /label= N-glycosylation_site
 XX
 PN WO9515171-A.
 XX
 PD 08-JUN-1995
 XX
 PF 05 DEC-1994; 94WO-US13967
 XX
 PP 03-DEC-1993; 930S-0163402.
 XX
 PA (CANC-) CANCER RES FUND CONTRA COSTA.
 XX
 PI Ceriani RL, Larocca DJ, Peterson JA;
 XX
 PP WPI: 1995-215151/28
 DR N-PSDB; AAQ91199.
 XX
 PT 46 kD apparent molecular weight human milk fat globule antigen
 PT used in assays to determine the presence of a cancerous tumour of
 PT epithelial origin, and in a vaccine against neoplastic tumours
 XX
 PS Example 7; Page 38-40; 68pp; English.
 XX
 CC A partial cDNA clone B446-1 (AAQ91199) for the 46 kDa HMFG antigen,
 CC a major component of the apical surface of the normal breast
 CC epithelial cell, was obtained by PCR of cDNA from a lactating
 CC breast cDNA library.
 XX
 SQ Sequence 218 AA;

Query Match 100.0%; Score 43; DB 16; Length 218;
 Best Local Similarity 100.0%; Pred No. 0.34;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VQFVASYKV 9
 DB 144 vqfvasykv 152
 ||| |||||

RESULT 5
 AAR77252
 ID AAR77252 standard; Protein; 387 AA.
 XX
 AC AAR77252;
 XX
 DT 21-NOV-1995 (first entry)
 XX
 DE HMFG 46 kDa antigen.
 XX
 KW HMFG; human milk fat globule; antigen; immunogen; vaccine; virucide;
 KW epithelium; tumor; breast cancer; monoclonal antibody; MAb.
 XX
 OS Homo sapiens.
 XX
 PN WO9515171-A.
 XX
 PD 08-JUN-1995.
 XX
 PF 05-DEC-1994; 94WO-US13967.
 XX
 PP Example 1993; 930S-0163402.
 XX
 PA (CANC-) CANCER RES FUND CONTRA COSTA.
 XX
 PI Ceriani RL, Larocca DJ, Peterson JA;
 XX
 PP WPI: 1995-215151/28
 DR N-PSDB; AAQ91199.
 XX
 PT 46 kD apparent molecular weight human milk fat globule antigen
 PT used in assays to determine the presence of a cancerous tumour of
 PT epithelial origin, and in a vaccine against neoplastic tumours
 XX
 PS Claim 6; Page 46-47; 68pp; English.
 XX
 CC A complete cDNA sequence for the 46 kDa HMFG antigen, a major
 CC component of the apical surface of the normal breast epithelial
 CC cell, was obtained by PCR and RACE methods. cDNA clones can be
 CC used to prepare MAb's for use in immunotherapy, immunodiagnosis,
 CC prognosis, diagnosis, imaging and therapy. Recombinant antigen can
 CC be expressed in prokaryotic or (glycosylated) in eukaryotic cells.

QY 1 VQFVASYKV 9
 DB 313 vqfvasykv 321
 ||| |||||

RESULT 6
 AAY94453
 ID AAY94453 standard; Protein; 387 AA.
 XX
 AC AAY94453;
 XX
 DT 11-SEP-2000 (first entry)
 XX
 DE Human lactadherin protein.
 XX
 KW Human; lactadherin; M5P-E8; anti-tumour; immune response;

Query Match 100.0%; Score 43; DB 16; Length 387;
 Best Local Similarity 100.0%; Pred No. 0.31;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VQFVASYKV 9
 DB 313 vqfvasykv 321
 ||| |||||

RESULT 6
 AAY94453
 ID AAY94453 standard; Protein; 387 AA.
 XX
 AC AAY94453;
 XX
 DT 11-SEP-2000 (first entry)
 XX
 DE Human lactadherin protein.
 XX
 KW Human; lactadherin; M5P-E8; anti-tumour; immune response;

KW exosome; dendritic cell.
 XX Homo sapiens.
 OS
 XX
 FH Key location/qualifiers
 FT Peptide 1..23
 FT /label Secretion_signal
 FT Protein 24..387
 FT /label lactadherin
 FT Binding-site 45..48
 FT /label Intracrin_binding_site
 XX
 XX EP1004664-A1.
 XX
 XX 31-MAY-2000.
 XX
 XX 24-NOV-1998; 98EP-0407925.
 XX
 XX 24-NOV-1998; 98EP-0407925.
 XX
 XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA (CORT-) INST CURTIE.
 PA
 XX WPI: 2000-452597/31.
 DR N F5DB, AAA27140.
 XX
 XX Chimeric isolated (human) lactadherin polypeptide that functions as an
 PT adaptor of cross-priming to eliminate pathogenic antigens, e.g. in
 PT tumors.
 PT
 XX Example 4; Page 12; 20pp; English.
 PS
 XX Lactadherin protein was found in exosomes produced by dendritic cells.
 XX The protein is involved in the phagocytosis of particulate antigens by
 CC dendritic cells. Exosomes produced by dendritic cells exposed to
 CC tumour antigens induce potent immune responses. Lactadherin or variants
 CC of it may be used in the mediation of an immune response. Variants of
 CC lactadherin may be used for inhibition and/or stimulation of the
 CC cross priming of antigens and stimulation of the phagocytosis of
 CC antigens by dendritic cells. Compositions derived from lactadherin can
 CC also be used to monitor an immune response, more specifically a CTL
 CC (cytotoxic T-lymphocyte) response and also to produce CTLs specific
 CC for a selected antigen. The present sequence is the human
 CC lactadherin protein.
 XX
 XX Sequence 387 AA;
 SQ
 Query Match 100.0%; Score 43; DB 21; Length 387;
 Best Local Similarity 100.0%; Pred. No. 0.61;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VQFVASYKV 9
 DB 313 VQFVASYKV 321
 |||||
 RESULT 7
 ABB44260
 ID ABB44260 standard; Peptide: 52 AA.
 XX
 AC ABB44260;
 XX
 XX 04-FEB-2002 (first entry)
 DT
 DE Peptide #11766 encoded by human foetal liver single exon probe.
 XX
 XX Human; foetal liver; gene expression, single exon nucleic acid probe.
 KW
 XX Homo sapiens.
 OS
 XX
 XX W-200157277 A2.
 PR

PD 09-AUG-2001.
 XX
 XX 30-JAN-2001; 2001WO-US00669.
 PF
 XX 04-FEB-2000; 2000US-0180312.
 FE 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PP 03-AUG-2000; 2000US-0632366.
 PP 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PP 04-OCT-2000; 2000GP-0034263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 FT
 XX WPI: 2001-484447/52.
 PP
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human fetal liver.
 PT
 XX Claim 27; SEQ ID NO 36895; 639pp + sequence listing; English.
 PS
 XX The invention relates to a single-exon nucleic acid probe for
 CC measuring human gene expression in a sample derived from human foetal
 CC liver. The single exon nucleic acid probes may be used for predicting,
 CC measuring and displaying gene expression in samples derived from human
 CC foetal liver. The present sequence is a peptide encoded by a single exon
 CC nucleic acid probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 XX
 XX Sequence 52 AA;
 SQ
 Query Match 83.7%; Score 36; DB 22; Length 52;
 Best Local Similarity 77.8%; Pred. No. 2.2;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 VQFVASYKV 9
 DB 23 VQFVASYKV 31
 |||||
 RESULT 8
 ABB27131
 ID ABB27131 standard; Protein: 52 AA.
 XX
 AC ABB27131;
 XX
 XX 23-JAN-2002 (first entry)
 DT
 DE Protein #9130 encoded by probe for measuring heart cell gene expression.
 XX
 XX Human; gene expression, heart, microarray, vascular system;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease.
 KW
 XX Homo sapiens.
 OS
 XX W-200157274-A2.
 PN
 XX 09-AUG-2001.
 PD
 XX 30-JAN-2001; 2001WO-US00666.
 PF
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PP 30-JUN-2000; 2000US-0608408.
 PP 03-AUG-2000; 2000US-0632366.
 PP 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PP

PR 04-OCT-2000; 2000GB-0024363.
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DP,
 XX WPI; 2001 483447/52.
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts -
 XX Claim 15. SEQ ID No. 28501, 550bp. English.
 XX The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart (see
 CC AB221335 AB221305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying
 CC gene expression in samples derived from the human heart via microarrays.
 CC By measuring gene expression, the probes are useful for prediction,
 CC diagnosis, grading, staging, monitoring and prognosing diseases of the
 CC human heart and vascular system e.g. cardiovascular disease,
 CC hypertension, cardiac arrhythmias and congenital heart disease.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX Sequence 52 AA;

Query Match 83.7%; Score 36; DB 22; Length 52;
 Best Local Similarity 77.8%; Pred. No. 2.2;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VQFVASVKV 9
 DB 23 vqfvgstkl 31

RESULT 9
 AAM58323
 ID AAM58323 standard; Protein: 52 AA.

XX AAM58323;
 XX 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe encoded protein SEQ ID No: 30428.
 XX Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer.

OS Homo sapiens

XX WO200157275-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001WO-0509967.

XX 04-FEB-2000; 2000US-0180312.

XX 25-MAY-2000; 2000US-0207456.

PP 30-JUN-2000; 2000US-0608408.

PP 04-AUG-2000; 2000US-0632466.

PP 21-SEP-2000; 2000US-0234687.

PP 27-SEP-2000; 2000US-0236459.

PP 04-OCT-2000; 2000GB-0024363.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DP;
 XX WPI; 2001 483447/52

XX Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 XX Example 4. SEQ ID No. 30428, 650bp + Sequence Listing; English.
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancer. The present sequence is a protein encoded by one of
 CC the probes of the invention.

XX Sequence 52 AA;

Query Match 83.7%; Score 26; DB 22; Length 52;
 Best Local Similarity 77.8%; Pred. No. 2.2;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VQFVASVKV 9
 DB 23 vqfvgstkl 31

RESULT 10
 AAM65298
 ID AAM65298 standard; Protein: 52 AA.

XX AAM65298;

XX 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe encoded protein SEQ ID No: 37403.
 XX Human; brain expressed exon; gene expression analysis; probe;
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
 KW epilepsy; cancer.

OS Homo sapiens.

XX WO200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-0509967.

XX 04-FEB-2000; 2000US-0180312.

XX 25-MAY-2000; 2000US-0207456.

PP 30-JUN-2000; 2000US-0608408.

PP 04-AUG-2000; 2000US-0632466.

PP 21-SEP-2000; 2000US-0234687.

PP 27-SEP-2000; 2000US-0236459.

PP 04-OCT-2000; 2000GB-0024363.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DP;
 XX WPI; 2001 483447/52.

XX Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -

XX Example 4. SEQ ID No. 37403, 650bp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancer. The present sequence is a protein encoded by one of

CC the probes of the invention.

XX Sequence 52 AA;

Query Match 83.7%; Score 36; DB 22; Length 52;

Best Local Similarity 77.8%; Pred. No. 2.2;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VQFVASYKV 9

DB 23 VQFVASYK 31

RESULT 11

AAM77997 standard; Protein; 52 AA.

XX AC

XX AAM77997;

DT 06-NOV-2001 (first entry)

XX Human bone marrow expressed probe encoded protein SEQ ID NO: 38303.

DE Human bone marrow expressed probe; gene expression analysis; probe;

KW microarray; cancer; leukaemia; lymphoma; myeloma.

XX Homo sapiens

XX W0200157276-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00668

XX 04-FEB-2000; 2000US-0180312

XX 26-MAY-2000; 2000US-0207456

XX 30-JUN-2000; 2000US-0608408

XX 03-AUG-2000; 2000US-0632366

XX 21-SEP-2000; 2000US-0234687

XX 27-SEP-2000; 2000US-0234687

XX 04-OCT-2000; 2000EP-0024263

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR,

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human bone marrow.

XX Example 4. SEQ ID NO. 38303, 658pp - Sequence listing; English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX bone marrow. They can be used to measure gene expression in bone marrow

XX samples which may enable the improved diagnosis and treatment of cancers

XX such as lymphoma, leukaemia and myeloma. The present sequence is a

XX protein encoded by one of the probes of the invention

XX Sequence 52 AA;

QY 1 VQFVASYKV 9

DB 23 VQFVASYK 31

Query Match 83.7%; Score 36; DB 22; Length 52;

Best Local Similarity 77.8%; Pred. No. 2.2;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0.

RESULT 12

AAM21889 standard; Protein; 52 AA.

XX AC

XX AAM21889;

DT 12-OCT-2001 (first entry)

XX Peptide #8123 encoded by probe for measuring cervical gene expression.

DE Probe; human; microarray; gene expression; cervical epithelial cell;

KW cervical cancer.

XX Homo sapiens.

XX W0200157278-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0234687.

XX 04-OCT-2000; 2000EP-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR,

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human cervical epithelial cells -

XX Claim 27. SEQ ID NO 26715; 487pp; English.

XX The present invention relates to human single exon nucleic acid probes

XX (SENPs; see AAM21889 AA128459). The present sequence is a peptide encoded

XX by one such probe. The SENPs are derived from human beta cells. The SENPs

XX can be used to produce a single exon microarray, which can be used for

XX measuring human gene expression in a sample derived from human cervical

XX epithelial cells. By measuring gene expression, the probes are therefore

XX useful in grading and/or staging of diseases of the cervix, notably

XX cervical cancer.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from Wipo

XX at http://wipo.int/pub/published_pat_sequences.

XX Sequence 52 AA;

Query Match 84.7%; Score 46; DB 22; Length 52;

Best Local Similarity 77.8%; Pred. No. 2.2;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VQFVASYKV 9

DB 23 VQFVASYK 31

RESULT 13

AAM38215

ID AAM38215 standard; Protein; 52 AA.

XX AC

XX AAM38215;

DT 17-OCT-2001 (first entry)

XX Peptide #12292 encoded by probe for measuring placental gene expression.

XX Probe: microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder.

XX Homo sapiens.

XX W0200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001W01500663

XX 04-FEB-2000; 20000S-0180312.

XX 26-MAY-2000; 2000S-0207456.

XX 30-JUN-2000; 2000S-0608408.

XX 03-AUG-2000; 2000S-0632966.

XX 21-SEP-2000; 2000S-0734687.

XX 27-SEP-2000; 2000S-0736359.

XX 04-OCT-2000; 2000S-0736359.

XX (MOLFE-) MOLECULAR DYNAMICS INC.

XX Penn St., Hanzel DR., Chen W., Park DE;

XX WPI; 2001-48887/53

XX Human genome derived single exon nucleic acid probes useful for

XX analyzing gene expression in human placenta

XX Claim 27; SEQ ID No 38484; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP;

XX see AAI31315-AA152546). The present sequence is a peptide encoded by one

XX such probe. The probes are useful for producing a microarray for

XX predicting, measuring and displaying gene expression in samples derived

XX from human placenta. The probes are useful for antenatal diagnosis of

XX human genetic disorders.

XX Sequence 52 AA;

RESULT 14

AAW4697

ID AAW4697 standard; Protein; 321 AA.

XX AAW4697;

XX 05-MAY-1999 (first entry)

XX Human milk fat globule protein MFG-E8.

XX Del-1; developmentally regulated endothelial cell factor.

XX discoidin 1; factor VII-like domain; epidermal growth factor; EGF;

XX diabetic retinopathy; rheumatoid arthritis; endometriosis;

XX angiogenesis.

XX Homo sapiens.

XX US5877281-A.

XX 02-MAP-1999.

XX 05-JUN-1996; 96US-0659235.

PR 05-JUN-1996; 96US-0659235.

PR 07-JUN-1995; 95US-0480229.

XX (PROG-) PROGENITOR INC.

XX (UYVA-) UNIV VANDERBILT.

XX Boston B., Overtemous T., Snodgrass DR., Zupancic TJ;

XX WPI; 1999-189720/16.

XX Del-1 polypeptide sequences useful for treatment of cancer,

XX diabetic retinopathy, rheumatoid arthritis and endometriosis

XX Example; Column 63-64; 7pp; English.

XX The present sequence represents human milk fat globule protein MFG-E8,

XX which has homology to the developmentally-regulated endothelial cell

XX factor 1 (Del-1). The Del-1 protein has epidermal growth factor (EGF) like

XX domain and discoidin 1/ factor VII-like domains. The Del-1 proteins

XX have an inhibitory effect on angiogenesis (blood vessel growth), this

XX activity may be useful clinically to prevent neovascularization of

XX tissues such as tumor nodules and prevention of metastases. The anti-

XX angiogenic activity of Del-1 may be used to treat abnormal conditions

XX that result from angiogenesis, including cancer, diabetic retinopathy,

XX rheumatoid arthritis and endometriosis. Since Del-1 promotes

XX angiogenesis it can be used to treat cardiac ischemia, thrombotic

XX stroke, wound healing and peripheral vascular disease. Del-1 is also

XX useful for promoting bone formation. Del-1 binds to alpha V beta 3

XX integrin, and is an apoptosis inducer.

XX Sequence 321 AA;

Query Match 83.7%; Score 46; DP 20; Length 421;

Best local Similarity 77.8%; Field No. 14;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VQFVASYKV 9

DB 246 VQIVASYKI 254

RESULT 15

AAW10364

ID AAW10364 standard; Protein; 480 AA.

XX AAW10364;

XX 03-MAY-1997 (first entry)

XX Mouse developmentally-regulated endothelial cell locus-1 protein.

XX Del-1; developmentally regulated endothelial cell locus-1;

XX signal transduction; cancer; tumor marker; angiogenesis;

XX diabetic retinopathy, rheumatoid arthritis, endometriosis;

XX cardiac ischemia; stroke; vascular disease; wound healing;

XX vulnerability; bone formation; diagnosis; therapy.

XX Mus sp.

XX 1.123

XX 24.480

XX 26.59

XX /note "epidermal growth factor-like domain 1"

XX 78.116

XX /note "epidermal growth factor-like domain 2"

XX 123.154

XX /label: EGF-3

```

FT      Domain      /note= "epidermal growth factor-like domain 3"
FT      158...314
FT      /label= Discoidin-1
FT      /note= "discoidin 1/factor VIII-like domain 1"
FT      319...476
FT      /label= Discoidin-2
FT      /note= "discoidin 1/factor VIII-like domain 2"
XX
XX
PN      WC9640769-A1.
XX
XX      19-DEC-1996.
XX
XX      05-JUN-1996; 96WO-US09456.
XX
XX      07-JUN-1996; 95DS-0480229.
XX
XX      (PROG-) PROGENITOR INC.
XX      (UYVA-) UNIV VANDERBILT.
XX
XX      Hoqan B, Quertermous T, Shodgrass HR, Zupancic TJ;
XX      WPI; 1997-052233/05.
XX      N-PSDB: AAI47338.
XX
XX      New developmentally regulated endothelial cell locus-1 (del 1) gene
XX      - used to develop prods. for the diagnosis and treatment of cancer
XX      and conditions involving abnormal angiogenesis
XX
XX      Claim 3; Fig 6; 137pp; English.
XX
XX      Murine Del-1 (AAW10364) is the polypeptide product of the murine
XX      developmentally-regulated endothelial cell locus-1 (del-1) gene
XX      (AA147338). It shows 94% amino acid homology to the human Del-1
XX      protein (AAW10365). Structurally, members of this novel gene family
XX      contain 3 EGF-like domains and 2 discoidin 1/factor VIII-like
XX      domains. Del 1 is expressed in endothelial and certain tumour
XX      cells. Its ability to inhibit vascular formation allows its used
XX      as an anti-angiogenic agent. It can be used as a tumour marker,
XX      to identify Del-1 binding partners, and to modulate endothelial
XX      cell growth and blood vessel formation. Recombinant Del-1 can be
XX      produced in transformed host cells utilising vectors incorporating
XX      del-1 nucleic acids.
XX
XX      Sequence 480 AA:

```

```

Query Match      83.7%; Score 36; DB 18; Length 480;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

OY      1 VQFVASVKV 9
        |||||
DB      402 VQFVASYK1 410

```

```

Search completed: September 5, 2002, 15:26:32
Job time: 372 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:28:53 ; Search time 69.78 Seconds
(without alignments)
12,993 Million cell updates/sec

Title: US-09-744-804-40
Perfect score: 43
Sequence: 1 VQFVASVKV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: PIP-71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	100.0	218	2	milc fat globule p
2	39	90.7	427	2	ags protein precu
3	36	83.7	409	2	PR47 protein - pi
4	34	79.1	463	1	milc fat globule m
5	33	76.7	401	2	glycoprotein anti
6	33	76.7	427	2	lysine--tRNA ligas
7	33	76.7	533	2	probable protein k
8	32	74.4	445	2	pol protein - simi
9	32	74.4	958	2	pol protein - simi
10	32	74.4	1161	2	probable lipid bind
11	31	72.1	213	2	hypothetical prote
12	31	72.1	237	2	hypothetical prote
13	31	72.1	308	2	glutamate--tRNA li
14	31	72.1	344	2	glutamate--tRNA li
15	31	72.1	489	1	hypothetical prote
16	31	72.1	1256	2	hypothetical prote
17	30	69.8	350	2	hypothetical prote
18	30	69.8	290	2	xyloglucan endogl
19	30	69.8	296	2	yeast translation
20	30	69.8	345	2	membrane bound ty
21	30	69.8	374	2	myosin-like i pla
22	30	69.8	396	2	JG2189
23	30	69.8	476	2	tubulin-interstitial
24	30	69.8	478	2	phospho beta-ala
25	30	69.8	529	2	hypothetical prote
26	30	69.8	576	2	hypothetical prote
27	30	69.8	641	2	DNA primase 00194
28	30	69.8	652	2	hypothetical prote
29	30	69.8	1215	2	spindle assembly

30	29	67.4	106	2	T17951	protein disulfide-
31	29	67.4	149	1	D1001D	discoilin 1 chain
32	29	67.4	149	2	B01822	discoilin 1 chain
33	29	67.4	187	2	AF1009	hypothetical prote
34	29	67.4	195	2	B9274	hypothetical prote
35	29	67.4	253	1	D1001A	discoilin 1 chain
36	29	67.4	253	1	D1001C	discoilin 1 chain
37	29	67.4	277	1	OR8337	cellular chitinidat
38	29	67.4	277	2	G91040	sulfate transport
39	29	67.4	277	2	A10665	sulfate transport
40	29	67.4	277	2	B85885	sulfate transport
41	29	67.4	277	2	A09811	sulfate transport
42	29	67.4	292	2	A47539	xyloglucan endo-1,
43	29	67.4	292	2	T10523	xyloglucan endo-1,
44	29	67.4	338	2	T28873	hypothetical prote
45	29	67.4	345	2	AC1757	yeast translation

ALIGNMENTS

RESULT 1
A47285
milc fat globule protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 28-May-1999
C:Accession: A47285
R:Baracca, D.; Petersen, J.A.; Berres, R.; Kuniyoshi, J.; Bistrain, A.M.; Geriani, R.L.
Cancer Res. 51, 4994-4998, 1991
A:Title: A M : 46,663 human milk fat globule protein that is highly expressed in huma
A:Reference number: A47285; M01D:91371351
A:Accession: A47285
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-218 <LARG>
A:Cross reference: GB:56151; R01:265996; PDB:AA19771.1; PDB:235397
C:Superfamily: milk fat globule protein; discoilin 1 and terminal hemolysis; EGF hom
F:1-57, Leu16, discoilin 1 amino terminal hemolysis; A:fragment 1-57
F:60-218/Domain; discoilin 1 amino terminal hemolysis <DN2>

Query Match 100.0% Score 43; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 0.067;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VQFVASVKV 9
DB 144 VQFVASVKV 152

RESULT 2
JC4915
ags protein precursor - rat
N:Alternate names: O-acetyl Gd3 ganglioside
C:Species: Rattus norvegicus (Norway rat)
C:Date: 28-Sep-1996 #sequence_revision 0; N-v-1996 #text_change 20 Jun 2000
C:Accession: JC4915
E:Gaur, K.; Nara, K.; Watanabe, Y.; Kohno, K.; Tai, T.; Sanui, Y.
Biochem. Biophys. Res. Commun. 225, 932-938, 1996
A:Title: Cloning and expression of cDNA for O acetylation of Gd3 ganglioside.
A:Reference number: JC4915; M01:4674422
A:Accession: JC4915
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-107 <ORF>
A:Cross reference: E01:384568; R01:462866; PDB:BAAL2121.1; PDB:4142007
A:Experimental source: GST cell
C:Comment: This protein is required for the O-acetylation of disialosganglioside siall
C:Genetics:
A:Gene: ags
C:Superfamily: milk fat globule protein; discoilin 1 amino terminal hemolysis; EGF hom
F:1-57, Leu16, signal sequence; A:status predicted -SD-
F:28-60/Domain; EGF homology <EG1>

F:68-107/Domain: EGF homology <E22>
 F:110-267/Domain: discoidin 1 amino terminal homology <E22>
 F:271-407/Domain: discoidin 1 amino terminal homology <E22>

Query Match 90.7% Score 39; DB 2; Length 427;
 Best Local Similarity 77.8%; Pred. No. 1;
 Matches 7; Conservative 2; Mismatches 0; Gaps 0.

QY 1 VQFVASVKV 9
 :|:|:|:|:|

DB 455 IOYVAAYKV 363

RESULT 4

T11743

PP47 protein - pi4 (fragment)

C:Species: Sus scrofa domestica (domestic pig)

C>Date: 16-Jul-1999 #sequence_revision 14-Jul-1999 #text_change 21 Jul 2000

C:Accession: T11743

R:Buslin, M.; Vogel, T.; Calvete, J. J.; Thole, H.H.; Schmidtke, J.; Matsuda, T.; Toepfer

Biol. Reprod. 58, 1957-1964, 1998

A:Title: Molecular cloning and characterization of P47, a novel bear sperm associated

A:Reference number: 217325; MUID:98296817

A:Accession: T11743

A>Status: preliminary; translated from ORF/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-404 <E22>

A:Cross-references: EMBL:Y11683; NID:q252927; PIR:CAA72379.1; PIR:q252928

A:Experimental source: testis

A:Function:

C:Description: may be involved in membrane remodeling and/or function as a zona pellucida

C:Superfamily: unassigned EGF-related proteins, EGF homology

F:6 407/Domain: EGF homology <E22>

Query Match 83.7% Score 36; DB 2; Length 409;
 Best Local Similarity 66.7%; Pred. No. 4.7;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0.

QY 1 VQFVASVKV 9

:|:|:|:|:|

DB 445 IOYVAAYKV 343

RESULT 4

A36479

Milk fat globule membrane protein - mouse

C:Species: Mus musculus (house mouse)

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: A36479

R:Stubbs, J.P.; Lokutis, C.; Sledge, K.L.; Bui, A.; Yuzhli, D.; Srinivasan, U.; Parry, C.

Proc. Natl. Acad. Sci. U.S.A. 87, 8417-8421, 1990

A:Title: cDNA cloning of a mouse mammary epithelial cell surface protein reveals the ex

A:Reference number: A36479; MUID:q104608

A:Accession: A36479

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-463 <E22>

A:Cross-references: EMBL:38337; NID:q109142; PIR:AAA25534.1; PIR:q109143

C:Superfamily: milk fat globule protein; discoidin 1 amino-terminal homology; EGF homolo

C:Keywords: membrane protein

F:29-60/Domain: EGF homology <E21>

F:68-107/Domain: EGF homology <E22>

F:147-404/Domain: discoidin 1 amino terminal homology <E22>

F:407-46/Domain: discoidin 1 amino terminal homology <E22>

Query Match

Best Local Similarity 79.1%; Score 34; DB 1; Length 463;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0.

QY 1 VQFVASVKV 9

BB 391 IOYVLSYKV 399
 :|:|:|:|:|

RESULT 5

S65138

glycoprotein antigen: MCP57/52, mammary gland - bovine (fragment)

R:Arnal, J.M.S.; Gysels, J.; Gysels, J.; Gysels, J.; Gysels, J.; Gysels, J.; Gysels, J.; Gysels, J.

C:Species: Bos primigenius taurus (cattle)

C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-Aug-1998

C:Accession: S65138; G48394

R:Aoki, N.; Kishi, M.; Taniguchi, Y.; Adachi, T.; Nakamura, R.; Matsuda, T.

Biochim. Biophys. Acta 1245, 385-391, 1995

A:Title: Molecular cloning of glycoprotein antigen MCP57/53 recognized by monoclonal

A:Reference number: S65138; MUID:96125736

A:Accession: S65138

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-401 <E22>

R:Mather, J.H.; Banghart, L.R.; Lane, W.S.

Biochem. Mol. Biol. Int. 29, 545-554, 1993

A:Title: The pig 147-glycoprotein, bovine mammary proteins, bovine mammary proteins

A:Reference number: A48394; MUID:93250576

A:Accession: G48394

A>Status: preliminary

A:Molecule type: protein

A:Residues: 207-220 <E22>

A:Experimental source: milk

A:Note: sequence extracted from NCBI backbone (NCBI:131457)

C:Superfamily: milk fat globule protein; discoidin 1 amino-terminal homology; EGF hom

C:Keywords: glycoprotein

F:1 32/Domain: EGF homology (fragment) <E21>

F:407/Domain: EGF homology <E22>

F:82-239/Domain: discoidin 1 amino-terminal homology <E21>

F:243 407/Domain: discoidin 1 amino terminal homology <E22>

Query Match 76.7% Score 33; DB 2; Length 401;

Best Local Similarity 55.6%; Pred. No. 21;

Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 VQFVASYKV 9

:|:|:|:|:|

DB 327 IOYVAAYKV 335

RESULT 6

S74211

PAS-6/7 protein precursor - bovine

C:Species: Bos primigenius taurus (cattle)

C>Date: 04-Dec-1997 #sequence_revision 12-Dec-1997 #text_change 20-Jan-2000

C:Accession: S74211; S78114; S24181

R:Barreard, J.; Andersen, M.H.; Berglund, L.; Kasmasen, J.T.; Peterson, T.E.

Eur. J. Biochem. 240, 628-636, 1996

A:Title: Characterization of glycoprotein PAS 6/7 from membranes of bovine milk fat

A:Reference number: S74211; MUID:97008954

A:Accession: S74211

A:Molecule type: mRNA

A:Residues: 1-427 <E22>

A:Cross-references: EMBL:X91895; MUID:91642779; PIR:CAA2997.1; PIR:q102779

A:Accession: S78114

A:Molecule type: protein

A:Residues: 1-427 <E22>

R:Kim, D.H.; Kameo, C.; Mizokami, Y.

Biochim. Biophys. Acta 1221, 203-211, 1992

A:Title: Purification and characterization of major glycoproteins, PAS 6 and PAS 7, 1

A:Reference number: S23926; MUID:92353107

A:Accession: S24181

A:Molecule type: protein

A:Residues: 383-394 <E22>

C:Superfamily: milk fat globule protein; discoidin 1 amino terminal homology; EGF hom

C:Keywords: blocked amino end; disulfide bond; glycoprotein; milk

R:Breton, R., Watson, D., Yabuuchi, M., Lepointe, J.
J Biol Chem 265:18248-18255, 1990
A:Title: Glutamine--tRNA synthetases of *Bacillus subtilis* 168T and of *Bacillus stearothermophilus*
A:Reference number: A36090; MIM:91009314
A:Accession: B36090
A:Molecule type: DNA
A:Residues: 1-489 <BRE>
A:Class: transferases; OR M55073; NIT: g143004; P104 AAA23434.1; P114 g143004; GK J05647
A:Note: the authors translated the codon AAT for residue 62 as Asp and CCG for residue 1
C:Genetics:
A:Gene: gltX
C:Superfamily: glutamate--tRNA ligase; glutamine--tRNA ligase homology
C:Keywords: aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis
F:5384/Domain: glutamine; tRNA ligase homology; EGL
F:256/Binding site: ATP (Lys) *status predicted

Query Match 72.18; Score 31; DB 1; Length 489;
Best local Similarity 62.58; Pred. No. 73;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VQFVASYK 8
:|||||
Db 262 IQFVASYK 269

Search completed: September 5, 2002, 15:28:54
Job time: 359 sec

GenCode version 4.5
Copyright (c) 1993 - 2000 Compton Ltd
OM protein - protein search, using sw model
Run on: September 5, 2002, 15:31:42 ; Search time 23.90 seconds
(without alignments)
10,252 Million cell updates/sec
Title: US-09-744-804-40
Perfect score: 43
Sequence: 1 VQFVASYKV 9
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224
Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.
SUMMARIES
Result No. Score Query % Match Length DB ID Description
1 43 100 0 387 1 MFCM_HUMAN Q08431 homo sapien
2 39 90.7 427 1 MFCM_RAT P78430 rattus norv
3 36 83.7 409 1 MFCM_PIG P79385 sus scrofa
4 34 79.1 463 1 MFCM_MOUSE P21956 mus muscu
5 33 76.7 427 1 MFCM_BOVIN Q05134 bos tauri
6 33 76.7 533 1 SVF_METMP Q30523 methanoco
7 32 74.4 445 1 MERI_SCHPO Q18252 schistosom
8 32 74.4 1161 1 POL_SEV1 P23074 simian foa
9 31 72.1 489 1 SVF_BACST P23249 bacillus st
10 30 69.8 641 1 PTM_OPEPA Q9926 ureaplasma
11 29 67.4 144 1 E3A_ECOLI P25688 escherich
12 29 67.4 223 1 HEX8_ADEP3 Q81473 porcine ad
13 29 67.4 253 1 E3A_ECOLI P25688 escherich
14 29 67.4 253 1 E3A_ECOLI P25688 escherich
15 29 67.4 277 1 CVST_ECOLI P16701 escherich
16 29 67.4 277 1 CVST_SALTY P11032 salmonella
17 29 67.4 351 1 FMI1_MOUSE Q08790 mus muscu
18 29 67.4 436 1 E3F1_BIFAM Q08431 elephas
19 29 67.4 485 1 STE_BACID Q08431 bacillus ha
20 29 67.4 560 1 TATP_NIVVF P17116 chrysidinae
21 29 67.4 560 1 TATP_NIVVF P17116 chrysidinae
22 29 67.4 928 1 RPT1_XPFLA P22114 oxytella p
23 29 67.4 1170 1 ITAL_HUMAN P20701 homo sapien
24 29 67.4 1477 1 YORI_YEAST P33349 saccharomy
25 29 67.4 2194 1 POL_BIE701 P25537 human enter
26 28 65.1 130 1 UCP6_SCHMA Q08431 schistosom
27 28 65.1 159 1 CA2_CAEEL Q08431 caenorhabdi
28 28 65.1 192 1 PAS2_DPOME P04388 drosophila
29 28 65.1 204 1 PAS2_HUMAN P04388 drosophila
30 28 65.1 207 1 HUS5_AQUAE Q08431 homo sapien
31 28 65.1 218 1 PAS5_HUMAN P03301 homo sapien
32 28 65.1 218 1 PAS5_MOUSE P10833 mus muscu
33 28 65.1 309 1 PAS1_YEAST P11119 saccharomy

ALIGNMENTS
RESULT 1
MFCM_HUMAN
ID MFCM_HUMAN STANDARD; PRT; 387 AA.
AC Q08431;
DT 01-OCT-1996 (Ref. 34, Created)
DT 01-NOV-1997 (Ref. 35, Last sequence update)
DT 16-OCT-2001 (Ref. 40, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-EG) (HMEG)
DE (breast epithelial antigen BA46) (MFCM) [Contains: Medin].
GN MFCM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast, and Breast carcinoma;
RX MEDLINE=96213608; PubMed=8639264;
PA Couto J.P., Taylor M.P., Godwin S.G., Ceriani P.L., Peterson J.A.;
RT "Cloning and sequence analysis of human breast epithelial antigen
RT BA46 reveals an RGD cell adhesion sequence presented on an epidermal
RT growth factor-like domain.";
RL DNA Cell Biol. 15:281-286(1996).
RN [2]
RP SEQUENCE OF 170-387 FROM N.A.
RX TISSUE=Mammary gland;
RX MEDLINE=91371351; PubMed=1909932;
PA Larocca P., Peterson J.A., Harro P., Kuniyoshi J., Bistrain A.M.,
RA Ceriani P.L.;
RT "A Mr 16,000 human milk fat globule protein that is highly expressed
RT in human breast tumors contains factor VIII-like domains.";
RL Cancer Res. 51:4994-4998(1991).
RN [3]
RP LACTADHERIN, AN RGD-DEPENDENT, ANTI-CHLAMYDIAL ACT.
RC TISSUE=Milk;
RX MEDLINE=9614924, PubMed 8559276;
PA Gierfida M.D., Cavallotto M., Giunta C., Conti A.,
RA Godovac-Zimmermann J.;
RT "Isolation and characterization of full and truncated forms of human
RT breast carcinoma protein BA46 from human milk fat globule membranes.";
RL J. Protein Chem. 17:143-148(1998).
RN [4]
RP SEQUENCE OF 259-317, AND IDENTIFICATION OF MFCM.
RX MEDLINE 99342076, PubMed 10411933;
PA Juchacz B., Ruckl J., Schuler P., Westmark G., Westmark G.,
RA Juchacz B., Ruckl J., Schuler P., Westmark G., Westmark G.;
RT "Medin, an integral fragment of active smooth muscle cell produced
RT Lactadherin forms the most common human cytokine.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:8634-8637(1999).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE 941688, PubMed 949329,
PA Taylor M.P., Couto J.P., Scallan C.D., Ceriani P.L., Peterson J.A.;
RT "Lactadherin (formerly BA46), a membrane-associated glycoprotein
RT expressed in human milk and breast carcinomas, promotes A13 Gly Asp
RT (RGD)-dependent cell adhesion.";

RL DNA Cell Biol 16:861-869(1997)
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
 CC SPECIFICALLY TO F-7-TRANS AND INHIBITS ITS REPLICATION.
 CC -1- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AOPTIC MEDIAL AMYLOID.
 CC -1- SURCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AOPTIC
 CC MEDIA. OVEREXPRESSED IN SEVERAL CAPCINOMAS.
 CC -1- PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
 CC AMINO ACID 264 AND 273.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN
 CC -1- SIMILARITY: CONTAINS 2 F5/R TYPE C DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/annotation/>
 CC or send an email to license@isb-sib.ch)

DR EMBL: U58516; AAC50549.1; -
 DR EMBL: S56151; AAC19771.1; -
 DR MIM: 602281; -
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR003421; FASRC_1.
 DR Pfam: PF00508; EGF_1.
 DR Pfam: PF00754; F5_P8_type_C; 2.
 DR SMART: SM00181; EGF_1.
 DR SMART: SM00231; FASRC_2.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01285; FASRC_1; 2.
 DR PROSITE: PS01286; FASRC_2; 2.
 KW Signal. Glycoprotein; Milk, secret. EGF-like domain. Any lipid.
 FT SIGNAL. 1 23
 FT CHAIN. 24 387 LACTADHERIN.
 FT CHAIN. 202 387 LACTADHERIN. SHORT FORM.
 FT CHAIN. 268 317 MEDIN.
 FT DOMAIN. 24 67 EGF-LIKE.
 FT DOMAIN. 70 225 F5/R TYPE C 1.
 FT DOMAIN. 230 387 F5/R TYPE C 2.
 FT SITE. 46 48 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID. 27 38 BY SIMILARITY.
 FT DISULFID. 32 55 BY SIMILARITY.
 FT DISULFID. 57 66 BY SIMILARITY.
 FT DISULFID. 70 225 BY SIMILARITY.
 FT DISULFID. 212 216 BY SIMILARITY.
 FT DISULFID. 230 387 BY SIMILARITY.
 FT CARBOHYD. 238 238 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD. 325 325 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD. 329 329 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD. 350 350 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE. 387 AA; 43123 MW; 2EE5571DEC837826 CRC64;

Query Match 100.0%; Score 43; DR 1; Length 387;
 Best Local Similarity 100.0%; Pred. No. 0.018;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VOFVASYKV 9
 |||||
 DB 313 VOFVASYKV 321

RESULT 2
 MFCM_RAT STANDARD. PRT. 427 AA
 AC P70450;
 DT 01-NOV-1997 (rel. 35, Created)
 DT 01-NOV-1997 (rel. 35, Last sequence update)
 DT 30-MAY-2000 (rel. 39, Last annotation update)
 DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (O-
 acetyl GD3 ganglioside synthase) (AGS) (MEGM).

GN MEGE8 OR AGS.
 OS Pampus norvegicus (Pai).
 CC Eukaryota; Metazoa; Chordata; Cephalo; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Ratius.
 OX NCBI_TaxID=10116;
 PN [1]
 PP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=76374422; PubMed=8780713;
 RA Ogura K., Nara K., Watanabe Y., Kohno K., Tai T., Sasaki Y.;
 FT "Cloning and expression of cDNA for O-acetylation of GD3
 RT ganglioside";
 FL Biochem Biophys Res Commun 225:932-938(1996).
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. SEEMS TO
 CC PARTICIPATE IN THE O-ACETYLAION OF GD3 GANGLIOSIDE SIALIC ACID
 CC -1- SURCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: SPLEEN, LUNG, HEART, BRAIN AND MUSCLE.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/R TYPE C DOMAINS
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/annotation/>
 CC or send an email to license@isb-sib.ch)

DR EMBL: P81058; PAA12210.1; -
 DR HSSP: P00740; IIXA.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001438; EGF_II.
 DR InterPro: IPR000421; FASRC_1.
 DR Pfam: PF00754; EGF_1.
 DR Pfam: PF00754; F5_P8_type_C; 2.
 DR PRINTS: P00010; EGF_2.
 DR SMART: SM00181; EGF_2.
 DR SMART: SM00231; FASRC_2.
 DR PROSITE: PS00022; EGF_1; 2.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS01285; FASRC_1; 2.
 DR PROSITE: PS01286; FASRC_2; 2.
 KW Signal; Glycoprotein; Repeat; EGF-like domain; Milk.
 FT SIGNAL. 1 22 POTENTIAL.
 FT CHAIN. 23 427 LACTADHERIN.
 FT DOMAIN. 24 61 EGF-LIKE 1.
 FT DOMAIN. 64 108 EGF-LIKE 2.
 FT DOMAIN. 111 267 F5/R TYPE C 1.
 FT DOMAIN. 272 427 F5/R TYPE C 2.
 FT DISULFID. 28 39 BY SIMILARITY.
 FT DISULFID. 44 44 BY SIMILARITY.
 FT DISULFID. 51 50 BY SIMILARITY.
 FT DISULFID. 68 79 BY SIMILARITY.
 FT DISULFID. 73 96 BY SIMILARITY.
 FT DISULFID. 98 107 BY SIMILARITY.
 FT DISULFID. 111 267 BY SIMILARITY.
 FT DISULFID. 254 258 BY SIMILARITY.
 FT DISULFID. 272 427 BY SIMILARITY.
 FT SITE. 87 89 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD. 61 61 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD. 230 230 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD. 280 280 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD. 390 390 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE. 427 AA; 47413 MW; EA8C8631F3EE6047 CRC64;

Query Match 100.0%; Score 39; DR 1; Length 427;
 Best Local Similarity 77.8%; Pred. No. 0.18;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 VOFVASYKV 9
 |||||
 DB 355 IQVASYKV 363

RESULT 3
MEGM_PIG
ID MEGM_PIG STANDARD; PRT: 409 AA.
AC P79385;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Lactadherin (Milk fat globule-EGF factor 8) (MFG E8) (MEGM) (Sperm surface protein SP47) (PP47).
GN MFGP8.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi.
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
OX NCBI_TaxID:9823;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23 35.
RC TISSUE=Testis;
RA Fossliin M.A.;
PL Submitted (NOV 1997) to the EMBL/GenBank/DBJ databases
CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
CC BINDING PROTEIN.
CC -1- SURFCELLULAR LOCATION: PERIPLASMIC MEMBRANE PROTEIN
CC TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC SPERMATOZOAN.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Gaps by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Y11683; CAA72379.1; -
CC HSPSP: P00740; LEDM
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR000421; FA58_C.
CC Pfam: PF00308; EGF; 2.
CC Pfam: PF00754; F5_F8_Type_C; 2.
CC SMART: SM00231; FA58C; 2.
CC PROSITE: PS01286; EGF_1; 2.
CC PROSITE: PS01286; EGF_2; 2.
CC PROSITE: PS01286; FA58C_1; 2.
CC PROSITE: PS01286; FA58C_2; 2.
CC GlycoProfile: Repeat, EGF-like domain.
FT DOMAIN 2 41 EGF-LIKE 1.
FT DOMAIN 44 98 EGF-LIKE 2.
FT DOMAIN 91 247 F5/8 TYPE C 1.
FT DOMAIN 252 409 F5/8 TYPE C 2.
FT SITE 67 69 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 6 17 BY SIMILARITY.
FT DISULFID 11 29 BY SIMILARITY.
FT DISULFID 31 40 BY SIMILARITY.
FT DISULFID 91 247 BY SIMILARITY.
FT DISULFID 234 238 BY SIMILARITY.
FT DISULFID 252 409 BY SIMILARITY.
FT CARBOHYD 41 41 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 372 372 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 409 AA, 45725 MW, 80507AE6029727A, -0.64.

Query Match
Best local Similarity 66 78; Score 35, Hs 1, Length 405,
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0.
QY 1 VQFVASYKV 9
Db 335 IQVAAVKV 343

RESULT 4
MEGM_MOUSE
ID MEGM_MOUSE STANDARD; PRT: 463 AA.
AC P21956; P97800;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-OCT-2001 (Rel. 40, Last annotation update)
DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (MEGM)
DE (Sperm surface protein SP47) (MP47).
GN MFGP8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23 35.
RC TISSUE=Mammary gland;
RA MEDLINE-91046308; PubMed-2122452;
PA Stubbs J.D., Lokatis C., Singer F.L., Bai A., Yuzaki D.,
KA Srivivasan U., Barry G.;
RT "cDNA cloning of a mouse mammary epithelial cell surface protein
RT reveals the existence of a distinct growth factor like domain linked
RT to factor VIII-like sequences";
PL Proc. Natl. Acad. Sci. U.S.A. 87:8417-8421(1990).
RP [2]
RP SEQUENCE OF 23-463 FROM N.A.
RC TISSUE=Testis;
RA Basslin M.A.;
PL Submitted (NOV 1997) to the EMBL/GenBank/DBJ databases
CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. ZONA PELLUCIDA-
CC BINDING PROTEIN.
CC -1- SURFCELLULAR LOCATION: PERIPLASMIC MEMBRANE PROTEIN.
CC TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND
CC SPERMATOZOAN.
CC -1- DEVELOPMENTAL STAGE: MPNA EXPRESSION IS DETECTABLE IN MAMMARY
CC TISSUE FROM NONPREGNANT ANIMALS & MAXIMAL IN THE LACTATING GLAND.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS
CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Gaps by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M48337; AAA39534.1; -
CC EMBL: Y11684; CAA72380.1; -
CC PIR: A36479; A36479.
CC HSPSP: P00740; LEDM.
CC Add, MG:102768; Mfgp8.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR001438; EGF_II.
CC InterPro: IPR000421; FA58_C.
CC Pfam: PF00308; EGF; 2.
CC Pfam: PF00754; F5_F8_Type_C; 2.
CC KATINLS, PR00010; EGF500D.
CC SMART: SM00181; EGF; 2.
CC SMART: SM00231; FA58C; 2.
CC PROSITE: PS00307; EGF_1; 2.
CC PROSITE: PS01196; EGF_2; 2.
CC PROSITE: PS01286; FA58C_1; 2.
CC PROSITE: PS01286; FA58C_2; 2.
CC Signal, GlycoProfile, Repeat, EGF-like domain, Milk.
FT SIGNAL 1 22
FT CHAIN 23 463 LACTADHERIN.
FT DOMAIN 24 61 EGF-LIKE 1.
FT DOMAIN 64 108 EGF-LIKE 2.
FT DOMAIN 148 303 F5/8 TYPE C 1.
FT DOMAIN 308 463 F5/8 TYPE C 2.

FT SITE 87 89 CELL ATTACHMENT SITE (POTENTIAL)
 FT DISULFID 28 39 BY SIMILARITY.
 FT DISULFID 33 49 BY SIMILARITY.
 FT DISULFID 51 60 BY SIMILARITY.
 FT DISULFID 68 79 BY SIMILARITY.
 FT DISULFID 73 96 BY SIMILARITY.
 FT DISULFID 98 107 BY SIMILARITY.
 FT DISULFID 148 303 BY SIMILARITY.
 FT DISULFID 290 294 BY SIMILARITY.
 FT DISULFID 308 463 BY SIMILARITY.
 FT CARBOHYD 61 61 N-LINKED (GLCNAc) (POTENTIAL).
 FT CARBOHYD 266 266 N-1-UNFED (GLCNAc) (POTENTIAL).
 FT CARBOHYD 316 316 N-1-UNFED (GLCNAc) (POTENTIAL).
 FT CARBOHYD 426 426 N-LINKED (GLCNAc) (POTENTIAL).
 FT CONFLICT 30 30 S -> F (IN REF. 2).
 FT CONFLICT 35 35 N -> D (IN REF. 1; AA SEQUENCE).
 FT CONFLICT 110 147 ETNYNDGEMFTTAVTAVTPTDLSNNLSNR ->
 G (IN REF. 2).
 FT CONFLICT 168 168 Y -> S (IN REF. 2).
 FT CONFLICT 196 196 H -> I (IN REF. 2).
 FT CONFLICT 309 309 L -> S (IN REF. 2).
 FT CONFLICT 395 395 E -> A (IN REF. 2).
 SQ SEQUENCE 463 AA; 51465 MW; 07886666PRRA724D CPC64;

Query Match 79.18; Score 34; DP 1; Length 463;
 Best Local Similarity 66.78; Pred. No. 3;
 Matches 6; Conservation 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VOFVASYKV 9
 : : : : :
 DB 391 LOVVFYSKV 366

RESULT 5
 ID MFCM_BOVIN STANDARD. PRT. 427 AA.
 AC Q95114; Q27959; P79344;
 DT 01-NOV-1997 (Ref. 35, Created)
 DT 01-NOV-1997 (Ref. 35, Last sequence update)
 DT 16-OCT-2001 (Ref. 40, Last annotation update)
 DE Lactadherin precursor (Milk fat globule-EPF factor 8) (MFG-E8)
 DE (MFG57/53) (PAS-5/PAS-7 glycoprotein) (MFCM) (Sperm surface protein
 SP47) (BP47) (Components 15/16).
 OS MFG-E8.
 GN Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Claviata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A. PARTIAL SEQUENCE AND CARBOHYDRATE-LINKAGE SITES.
 RP STRAIN-BOLSTEIN; TISSUE:Mammary gland;
 RX MEDLINE=97008954; PubMed=8856064;
 RA Hvarregaard J., Andersen M.H., Berglund L., Rasmussen J.T.,
 RA Petersen T.E.;
 RA "Characterization of glycoprotein PAS-6/7 from membranes of bovine
 PT milk fat globules".
 PL Pur. J. Biochem. 240:628-636(1996)
 RN [2]
 RP SEQUENCE OF 18-427 FROM N.A.
 RP TISSUE=Mammary gland;
 RX MEDLINE=96125736; PubMed=8541316;
 RA Aoki N., Kishi M., Taniguchi Y., Adachi T., Nakamura R.,
 RA Matsuda T.;
 RA "Molecular cloning of glycoprotein MFG-57/53 recognized by
 PT monoclonal antibodies raised against bovine milk fat globule
 RT membranes".
 RL Biochim Biophys Acta 1245:385-391(1995)
 RN [3]
 RP SEQUENCE OF 19-427 FROM N.A.
 RP TISSUE=Testis;
 RC TISSUE=Testis;
 RA Ensslin M.A.;

Submitted (Nov 1997) to the EMBL/Genbank/DBT databases.
 [4]
 SEQUENCE OF 140-146; 174-187; 231-246 AND 422-427.
 TISSUE=Milk;
 RX MEDLINE=94258576; PubMed=8485476;
 RA Mather I.H., Banghart L.R., Lane W.S.;
 PT "The major fat-globule membrane proteins, bovine components 15/16 and
 PT galacta pig GP 55, are homologous to MFG-E8, a murine glycoprotein
 PT containing epidermal growth factor-like and lactoferrin-like
 PT sequences".
 PL Biochem. Mol. Biol. Int. 29:545-554(1993).
 CC FUNCTION: PROBABLY ASSOCIATES WITH PHOSPHOLIPIDS ON THE SURFACE OF
 CC MAMMARY EPITHELIAL CELLS AND MILK FAT GLOBULES. ZONA PELLUCIDA-
 CC BINDING PROTEIN.
 CC 1- ALTERNATIVE PROTEIN: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING THE SHORT FORM
 CC LACKS 53 AMINO ACIDS WITHIN THE F578 TYPE C1 DOMAIN.
 CC 2- TISSUE SPECIFICITY: MILK AND SPERMATOZOAN.
 CC 3- FIVE FIVE 2 O-LINKED GLYCANS CONSIST OF GAL, GLCNAc AND FUC, WITH
 CC PROBABLY FIVE AS REDUCING TERMINAL SUGAR.
 CC 4- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC 5- SIMILARITY: CONTAINS 2 F578 TYPE C1 DOMAINS.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed (save by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/annouce>)
 CC or send an email to license@isb-sib.ch).

EMBL: X91895; FAA62997.1; -;
 EMBL: S80643; AAB35894.2; -;
 EMBL: Y11719; CAA72406.1; -;
 DSSP: P00740; 1EDM.
 InterPro: IPR000561; EGF-like.
 Pfam: PF00008; EGF_2.
 Pfam: PF00754; F5_F8_type_C; 2.
 SMART: SM00181; EGF_2.
 SMART: SM00231; PAS6C; 2.
 PROSITE: PS00622; EGF_1; 2
 PROSITE: PS01186; EGF_2; 2
 PROSITE: PS01285; PAS6C_1; 2.
 PROSITE: PS01286; PAS6C_2; 2.
 Signal: Glycoprotein, Milk, Repeat, EGF like domain;
 KW Alternative Splicing.
 FT SIGNAL 1 18
 FT CHAIN 19 427 LACTADHERIN.
 FT DOMAIN 20 59 EGF-LIKE 1.
 FT DOMAIN 62 106 EGF-LIKE 2.
 FT DOMAIN 109 265 F578 TYPE C 1.
 FT DOMAIN 270 427 F578 TYPE C 2.
 FT SITE 85 87 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 24 35 BY SIMILARITY.
 FT DISULFID 29 47 BY SIMILARITY.
 FT DISULFID 49 58 BY SIMILARITY.
 FT DISULFID 66 77 BY SIMILARITY.
 FT DISULFID 71 94 BY SIMILARITY.
 FT DISULFID 96 105 BY SIMILARITY.
 FT DISULFID 109 265
 FT DISULFID 252 256
 FT DISULFID 270 427
 FT CARBOHYD 27 27
 FT CARBOHYD 34 34
 FT CARBOHYD 54 54
 FT CARBOHYD 227 227
 FT CARBOHYD 169 221
 FT VARSPIC 19 19 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 28 28 A -> F (IN REF. 1).
 FT CONFLICT 28 28 I -> Q (IN REF. 1).
 SQ SEQUENCE 427 AA; 47411 MW; 4CBEE3A1DC4EB24 CRC64;

```

16-OCT-2001 (Rel. 40, last annotation update)
DE Metabolic specificity: Serine/threonine protein kinase-motif (Ref. 2,7,1,1,37).
DE MEK1 OR SPAC1464.03.
CC Sciencelife/charom/yes (Pisum yeast).
CC Eukaryota, Fungi, Ascomycota, Schizosaccharomycetes;
CC Schizosaccharomycetes, Schizosaccharomycetaceae;
CC Schizosaccharomycetes.
CC NBI_TaxID=4896;
CC [1]
RN SEQUENCE FROM N.A.
RN STRAIN=972;
RA Type M.H.; Bryant J.A., Aves S.J.;
RN Submitted (jul-1996) to the EMBL/Genbank/Joint databases.
CC [2]
RN SEQUENCE FROM N.A.
RA Protein K.; Charpentier C.M., Bartell R.G., Benjandream H.A., Wood V.;
CC Submitted (Aug-1997) to the EMBL/Genbank/Joint databases.
CC 1. FUNCTION: FORMABLE PROTEIN KINASE REQUIRING FOR METOLIC
CC RECOGNITION.
CC 2. CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphorylated protein.
CC 3. SIMILARITY: BELONGS TO THE SERINE/THREONINE KINASES.
CC 4. SIMILARITY: CONTAINS 1 FHA DOMAIN
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed (usage by and for commercial
CC entities requires a license agreement) (see http://www.ebi.ac.uk/submit/).
CC Cited as entry in: http://www.ebi.ac.uk/submit/
CC -----
CC PDBL: 274478; CAA96101.1; 1.
CC PDBL: 294594; CARI1196.1; 1.
CC BSSP: P00518; 198K.
CC Ilic-Pro: IP900719; PKK-kinase.
CC InterPro: IPR000253; FHA-domain.
CC InterPro: IPR002290; Ser_thr_kinase.
CC Pfam: PF00498; FHA; 1.
CC Pfam: PF00069; pkinase; 1.
CC SMART: SM00240; FHA; 1.
CC SMART: SM00240; SLK; 1.
CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
CC PROSITE: PS00008; FHA_DOMAIN; 1.
CC Transferrase, Serine/threonine protein kinase, ATP-binding; Metosis.
CC KW TRANSFERIN 62 116 FHA.
CC FT TRANSFERIN 160 421 PROTEIN KINASE.
CC FT NO_BIND 166 174 ATP (BY SIMILARITY).
CC AC1_S11H 281 281 RE SIMILARITY.
CC FT TRANSFERIN 160 166 ATP (BY SIMILARITY).
CC SEQUENCE 445 AA; 5.151 MW; 55297996AIP45B0 CRC64;
CC -----
CC Query Match: 74.4%; Score 42; DB 1; Length 445;
CC Best Local Similarity 75.0%; Pred. No. 8.5;
CC Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC -----
Q7 1 VQTVASYK 8
Db 75 1QTVASYK 82
CC -----
RESULT 8
POL_SFV1 POL_SFV1 STANDARD; PRT: 1161 AA.
ID POL_SFV1 AC P23074;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-JUL-1993 (Rel. 26, 1st sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE DE Polyprotein (contains: Protease {EC 3.4.23. }, Reverse
DE transcaprase {EC 3.7.7.40}, Endonuclease).

```

```

Query Match          76.7%   Score 33; DB 1; Length 427;
Best Local Similarity 85.6%; Prod. No. 4.7;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 LVVASYKV 9
        :|::||::|
DD      353 LYVAAYEV 361

RESULT 6
SYK_METMP
ID SYK_METMP STANDARD; PRT; 533 AA.
AC Q30522;
15-JUL-1998 (Ref. 36, Created)
15-JUL-1998 (Ref. 36, Last sequence update)
16-OCT-2001 (Ref. 46, Last annotation update)
Lysyl tRNA synthetase (EC 6.1.1.6) (Lysine tRNA ligase) (LYSRS).
DE LYS.
OS Methanococcus mariprofundus.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteraceae;
OC Methanococcus
OX NCBI_TaxID=39152;
[1]
SEQUENCE FROM N.A., AMP SEQUENCE OF 1-22.
STRAIN:JG.
MFLEINF 36016225 PubMed 940002
RA Ida M., Morgan S., Garman A.W., Richmond P., Withnough D.C.,
Gardner W., Liu W., Wilson R.F., Scott J.L.
FE "A euryarchaeal lysyl tRNA synthetase, resembling to class I
synthetases."
FT Science 278:1119-1122(1997).
CC -I- CATALYTIC ACTIVITY: AMP + L-Lysine + tRNA(Lys) -> AMP + diphosphate
+ L-Lysyl-tRNA(Lys).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- SIMILARITY: BELONGS TO CLASS-I AMINACYL-TRNA SYNTHETASE FAMILY.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed (except by and for commercial
entities requires a license agreement (See http://www.isb.sib.ch/announce/
or send an email to licensewiz@sib.ch)).
-----
EMBL; AF009824; AAB87416.1; .
InterPro; IP001412; tRNA-synt_1.
InterPro; IP003904; tRNA-synt_lysl.
Pfam; PF01921; tRNA-syn_L1; 1.
PFam-SITE; PS00178; AA_tRNA_LIGASL_1; 1.
Amino-acid; tRNA synthetases; Protein families; Ligase; ATP binding.
SITE 29 94 "HIGH" REGION.
SITE 278 282 "XMSKS" REGION.
CONFLICT 11 11 I -> L (in Ref. 1; AA SEQUENCE).
SEQUENCE 533 AA; 61373 MW; 607FA868A7AA602 CF664.

Query Match          76.7%   Score 33; DB 1; Length 533;
Best Local Similarity 85.7%; Prod. No. 6;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 FVASYKV 9
        ||::::|
DB      497 FVASYKI 503

RESULT 7
MEK1_SCHPO
ID MEK1_SCHPO STANDARD; PRT; 445 AA.
AC Q10242;
16-OCT-2001 (Ref. 46, Created)
16-OCT-2001 (Ref. 49, Last sequence update)

```

GN 1001
OS Simian foamy virus (type 1) (SPV-1).
OC Viruses, Retroviruses; Retroviridae; Spumavirus.
OX NCBI_TaxID=11964;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=91276270; PubMed=1647358;
RA Kiplov I., Kay A., Hayat M., Pavlov P., Porias J., Galibert F.;
RI "Sequence analysis of the simian foamy virus type 1 genome.";
RL Gene 101:185-194(1991).
RN 121
RP SEQUENCE OF 1-970 FROM N.A.
RX MEDLINE=91361538; PubMed=1653483;
RA Morgia A., Luciw P.A.;
RI "Replication and regulation of primate foamy viruses.";
RL Virology 184:475-482(1991).
RN 131
RP SEQUENCE OF 969 1161 FROM N.A.
RX MEDLINE=90801049; PubMed=2152825;
RA Morgia A., Shaw K.E.S., Jackson J.B., Luciw P.A.;
RI "Relationship of the gag genes and the nucleocapsid domain of the gag genes of simian foamy virus type 1 and human foamy virus";
RL J. Virol. 64:406-410(1990).
CC 1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A9.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
DR EMBL: X54482; NOT_ANNOTATED; GUS
DR EMBL: X58484; GAA41394.1; -
DR EMBL: M33561; AAA47793.1; -
DR PIR: A33562; A33562
DR PIR: S15566; S15566
DR PIR: S18738; S18738.
DR HSSP: P03355; IMML
DR MEMOPS: A09.001; -
DR InterPro: IPR001156; RNaseH
DR InterPro: IPR004477; RNaseH
DR InterPro: IPR001584; Rve.
DR InterPro: IPR001641; Spuma_Agptase.
DR Pfam: PF000075; RNaseH; 1
DR Pfam: PF00665; rvo; 1
DR Pfam: PF00078; rvt; 1
DR PRINTS: PK00920; SPUMVIRPTASE.
KW transferase; PRA directed DNA polymerase, Hydrolase; Nuclease;
KW Endonuclease; Aspartyl protease, Polypeptidase.
FT ACT SITE 36 36 BY SIMILARITY.
FT CONFLICT 236 236 T -> I (IN REF. 2).
FT CONFLICT 910 910 S -> G (IN REF. 2).
FT CONFLICT 950 950 A -> T (IN REF. 2).
SQ SEQUENCE 1161 AA; 131717 MW; 06D09D1DBB4C14C CRC64;

Query Match 74 49 Score 32; DB 1; Length 1161;
Res' local Similarity 65.64; Pred No 16;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0.

QY 1 VOFVASYK 9
ID SYE_BACST STANDARD; PPT; 489 AA
AC P22246;
DT 01-AUG-1991 (Rel. 19, Created)

GN 1001
OS Simian foamy virus (type 1) (SPV-1).
OC Viruses, Retroviruses; Retroviridae; Spumavirus.
OX NCBI_TaxID=11964;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=91276270; PubMed=1647358;
RA Kiplov I., Kay A., Hayat M., Pavlov P., Porias J., Galibert F.;
RI "Sequence analysis of the simian foamy virus type 1 genome.";
RL Gene 101:185-194(1991).
RN 121
RP SEQUENCE OF 1-970 FROM N.A.
RX MEDLINE=91361538; PubMed=1653483;
RA Morgia A., Luciw P.A.;
RI "Replication and regulation of primate foamy viruses.";
RL Virology 184:475-482(1991).
RN 131
RP SEQUENCE OF 969 1161 FROM N.A.
RX MEDLINE=90801049; PubMed=2152825;
RA Morgia A., Shaw K.E.S., Jackson J.B., Luciw P.A.;
RI "Relationship of the gag genes and the nucleocapsid domain of the gag genes of simian foamy virus type 1 and human foamy virus";
RL J. Virol. 64:406-410(1990).
CC 1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A9.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
DR EMBL: X54482; NOT_ANNOTATED; GUS
DR EMBL: X58484; GAA41394.1; -
DR EMBL: M33561; AAA47793.1; -
DR PIR: A33562; A33562
DR PIR: S15566; S15566
DR PIR: S18738; S18738.
DR HSSP: P03355; IMML
DR MEMOPS: A09.001; -
DR InterPro: IPR001156; RNaseH
DR InterPro: IPR004477; RNaseH
DR InterPro: IPR001584; Rve.
DR InterPro: IPR001641; Spuma_Agptase.
DR Pfam: PF000075; RNaseH; 1
DR Pfam: PF00665; rvo; 1
DR Pfam: PF00078; rvt; 1
DR PRINTS: PK00920; SPUMVIRPTASE.
KW transferase; PRA directed DNA polymerase, Hydrolase; Nuclease;
KW Endonuclease; Aspartyl protease, Polypeptidase.
FT ACT SITE 36 36 BY SIMILARITY.
FT CONFLICT 236 236 T -> I (IN REF. 2).
FT CONFLICT 910 910 S -> G (IN REF. 2).
FT CONFLICT 950 950 A -> T (IN REF. 2).
SQ SEQUENCE 1161 AA; 131717 MW; 06D09D1DBB4C14C CRC64;

Query Match 74 49 Score 32; DB 1; Length 1161;
Res' local Similarity 65.64; Pred No 16;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0.

QY 1 VOFVASYK 9
ID SYE_BACST STANDARD; PPT; 489 AA
AC P22246;
DT 01-AUG-1991 (Rel. 19, Created)

DT 01-AUG-1991 (Rel. 19, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glutamyl-tRNA synthetase (Eukaryotic) (RNA liqase) (GluRS).
GN GLTX.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Geobacillus.
OX NCBI_TaxID=1422;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=91309314; PubMed=2190236;
RA Rerton P., Watson D., Yaguchi M., Lapointe J.;
RI "Glutamyl-tRNA synthetases of Bacillus subtilis 168" and of Bacillus stearothermophilus: cloning and sequencing of the gltX genes and comparison with other aminoacyl-tRNA synthetases.";
FL J Biol Chem. 265:18249-18255(1990)
CC 1- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) -> AMP + diphosphate + L-glutamyl-tRNA(Glu).
CC 1- SUBUNIT: MONOMER.
CC 1- SUBCELLULAR LOCATION: Cytoplasmic
CC 1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
DR EMBL: M55072; AAA22494.1; -
DR PIR: K30900; SYRSES
DR HSSP: P27000; IGLN
DR InterPro: IPR000934; tRNA-synt_1c
DR InterPro: IPR001412; tRNA-synt_1c; 1
DR Pfam: PF00749; tRNA-synt_1c; 1
DR PRINTS: PF00987; tRNA-syntHGLU.
DR PROSITE: PS00178; AA-tRNA_LIGASE_I; 1
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT SITE 11 21 "HIGH" REGION.
FT SITE 253 257 "KMSKS" REGION.
FT BINDING 256 256 ATP (BY SIMILARITY).
SQ SEQUENCE 489 AA; 5418 MW; A6126E6842F154C64;

Query Match 72 19; Score 31; DB 1; Length 489;
Res' local Similarity 62.58; Pred No 16;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0.

QY 1 VOFVASYK 8
ID PRIM_HPEPA STANDARD; PPT; 641 AA
AC Q9PPZ6;
DT 16-OCT-2001 (Rel. 40, Created)
DI 16-OCT-2001 (Rel. 40, Last sequence update)
DE DNA primase (EC 2.7.7.7).
GN DNAG OR U0494.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma
OX NCBI_TaxID=134821;
RN 11
RP SEQUENCE FROM N.A.
AC STPATN=SPF0VAP 3;
FX MFDI INE=20500219; PubMed=11048724;
DT 01-AUG-1991 (Rel. 19, Created)

```

RA Cassell G.H.:
RT "the complete sequence of the mucosal pathogen Ureaplasma
RL urealyticum".
CC -1- FUNCTION: DNA PRIMASE IS THE POLYMERASE THAT SYNTHESIZES SMALL
CC RNA PRIMERS FOR THE OKAZAKI FRAGMENTS ON BOTH TEMPLATE STRANDS AT
CC REPLICATION FORKS DURING CHROMOSOMAL DNA SYNTHESIS.
CC -1- COFACTOR: BINDS ONE ZINC ION PER MOLECULE (BY SIMILARITY).
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/submit-sib.ch)
CC or send an email to license@sib.ch
CC
CC EMBL: AF002147; AAF30906.1; -.
CC RSP: Q9X4D0; IP00.
CC InterPro: IPR002936; Toprim.
CC InterPro: IPR002694; Znf_CHCC.
CC Plam: PF01751; Toprim; 1.
CC Plam: PF01807; Zf-CHC2; 1.
CC Pfam: PF0002988; Znf_CHCC; 1.
CC SMART: SM00493; TOPRIM; 1.
CC SMART: SM00400; Znf_CHCC; 1.
CC Transferrase; DNA replication; DNA-directed RNA polymerase; Primosome;
CC Zinc finger, zinc, Metal binding, Complete protein.
CC ZNFING 39 63 CHC2-TYPE (BY SIMILARITY).
CC ZNFING 641 AA; 75390 MW; A6920160D58EF47 CRC64;
CC
CC Query Match 69.8%; Score 30; DB 1; Length 641;
CC Best Local Similarity 62.5%; Pred. No. 38;
CC Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
CC
QY 1 VQFVASYK 9
DE 70 KIVADYK 77
CC
CC RESULT 11
CC DISD.DICDI
CC ID DISD.DICDI STANDARD; PRT; 149 AA.
CC AC P02888;
CC DT 21-JUL-1986 (Rel. 01, Created)
CC DT 21-JUL-1986 (Rel. 01, Last sequence update)
CC DT 15-JUL-1998 (Rel. 36, Last annotation update)
CC DE Discolidin 1, D chain (Fragment).
CC GN DISC.
CC OS Dictyostelium discoideum (Slime mold).
CC EC Eukaryotic; Molecular; Cytoskeletal; First position
CC NCBI_TaxID:44689;
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE:82170475; PubMed:6279874;
CC Poole S., Firtel R.A., Lamar E., Rowekamp W.
CC "Sequence and expression of the discolidin 1 gene family in
CC Dictyostelium discoideum."
CC J. Mol. Biol. 153:273-286(1991)
CC [2]
CC CELL ATTACHMENT SITE.
CC MEDLINE:95074463; PubMed:6509559;
CC Springer W.R., Cooper D.N.W., Baronides S.H.;
CC "Discolidin 1 is implicated in cell-substratum attachment and ordered
CC cell migration of Dictyostelium discoideum and resembles
CC fibronectin."
CC Cell 39:557-564(1984).
CC -1- FUNCTION: GAI-A-TSE- AND N-A-TYI-PALATIN-SAMININ-BINDING 190-TIN. MAY
CC PLAY A ROLE IN CELL-SUBSTRATUM ADHESION RATHER THAN IN CELL-CELL
CC ADHESION. MAY BE NECESSARY FOR THE MAINTENANCE OF NORMAL ELONGATE
CC MORPHOLOGY DURING AGGREGATION.
CC
CC -1- SUBUNIT: TETRAMER OF FOUR DIFFERENT CHAINS (A TO D).
CC -1- SURCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: STALK CELLS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/submit-sib.ch)
CC or send an email to license@sib.ch
CC
CC EMBL: J01285; AAA33200.1; -.
CC PIR: A04583; DUBO1D.
CC Cite: Eb. E602002; discd.
CC InterPro: IPR000421; FA58_C;
CC Plam: PF00754; F5-F8_type_C; 1.
CC SMART: SM00231; FA58C_1;
CC DR PROSITE: PS01285; FA58C_1; 1.
CC DR PROSITE: PS01286; FA58C_2; FALSE_NEG.
CC KW Lectin; Cell adhesion.
CC FT SITE 79 81 CELL ATTACHMENT SITE.
CC FT NON_TER 149 149
CC SQ SEQUENCE 149 AA; 16746 MW; 1B8E0F223B249A1 CRC64;
CC
CC Query Match 67.4%; Score 29; DB 1; Length 149;
CC Best Local Similarity 62.5%; Pred. No. 13;
CC Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
CC
QY 2 QFVASYKV 9
DE 84 QWTSYKI 91
CC
CC RESULT 12
CC HEX8.ADEP3
CC ID HEX8.ADEP3 STANDARD; PRT; 223 AA.
CC AC Q83453; Q9YIR6;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Hexon-associated protein precursor (Protein VIII).
CC GN PVIII.
CC OS Porcine adenovirus type 3 (PAV-3).
CC CC Viruses; ds-DNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
CC OX NCBI_TaxID:35265;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=6618;
CC RX MEDLINE:95351046; PubMed:7625129;
CC Roddy P.S., Noy P., Derbyshire J.B.;
CC "Sequence analysis of poliovirus (VLP), E2 and E2-like regions of porcine
CC adenovirus type 3."
CC J. Virus Res. 36:97-106(1995).
CC [2]
CC SEQUENCE FROM N.A.
CC STRAIN=6618;
CC RX MEDLINE:99056141; PubMed:9837805;
CC Roddy P.S., Hamakuchi H., Song J.Y., Lee J.H., Ryan B.H., Park J.H.,
CC Cha S.H., Park Y.T., Yoon S.K., Park H.A.;
CC "Nucleotide sequence and transcription map of porcine adenovirus type
CC 3."
CC J. Virol. 251:414-426(1998).
CC [3]
CC SEQUENCE FROM N.A.
CC STRAIN=6618 / IAF;
CC RX Larocque D., Malenfant F., Massie R., Dea S.;
CC "Porcine adenovirus serotype 3, complete genome."
CC Submitted (Feb 1999) to the EMBL/Genbank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

DR EMBL: AF083132; AAC99446.1; -;
 DR EMBL: AJ237815; CAB41034.1; -;
 DR EMBL: AB024117; BAA76972.1; -;
 DR InterPro: IPR000446; Adenovirus_F1;
 DR Pfam: PF01410; Adenovirus_F1; 1;
 KW hexon-associated protein.
 FT PROPEP 1 111
 FT CHAIN 112 223
 FT SITE 111 112
 FT SITE 112 112
 SQ SEQUENCE 223 AA: 24261 MW: 7761342411060261 GRC64;
 BY SIMILARITY
 DE HEXON ASSOCIATED PROTEIN.
 DE C1FAVACP (BY ADENOVIRUS PROTEINASE).

Query Match 67.48; Score 29; DB 1; Length 223;
 Best Local Similarity 62.54; Pred. No. 23;
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 QFVASYKV 9
 I I I I I
 DB 207 QFIANYNI 214

RESULT 14
 DISC_DICDI
 ID DISC_DICDI STANDARD; PRT; 253 AA.
 AC P02886;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Discoidin I, A chain.
 GN DSCA.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=82170475; PubMed=6279874;
 RA Poole S., Firtel R.A., Lamar E., Rowekamp W.;
 RT "Sequence and expression of the discoidin I gene family in
 RT Dictyostelium discoideum."
 RL J. Mol. Biol. 153:274-289(1991).
 RN [2]
 RP REVISIONS, SEQUENCE FROM N.A.
 RX MEDLINE=85059644; PubMed=6754951;
 RA Jellingshaus U., Schaefer H., Schmidt W., Rowekamp W.;
 RT "Transcription of a dictyostelium discoidin-i gene in yeast
 RT alternative promoter sites used in two different eukaryotic cells."
 RL J. Mol. Biol. 150:633-636(1992).
 RN [4]
 RP SEQUENCE OF 1-40 FROM N.A.
 RX MEDLINE=9223698; PubMed=6284473;
 RA Devine J.M., Tsang A.S., Williams J.G.;
 RT "Differential expression of the members of the discoidin I multigene
 RT family during growth and development of Dictyostelium discoideum."
 RL Cell 28:793-800(1982).
 RN [4]
 RP CELL ATTACHMENT SITE.
 RX MEDLINE=85074463; PubMed=6509552.
 RA Springer W.R., Cooper D.N.W., Barondes S.H.;
 RT "Discoidin I is implicated in cell-substratum attachment and ordered
 RT cell migration of Dictyostelium discoideum and resembles
 RT fibronectin."
 RL Cell 39:557-564(1984).
 CC -1- FUNCTION: GALACTOSE- AND N-ACETYLGLACTOSAMINE-BINDING LECTIN. MAY
 CC PLAY A ROLE IN CELL-SUBSTRATUM ADHESION RATHER THAN IN CELL-CELL
 CC ADHESION. MAY BE NECESSARY FOR THE MAINTENANCE OF NORMAL ELONGATE
 CC MORPHOLOGY DURING AGGREGATION.
 CC -1- SUBUNIT: TETRAMER OF FOUR DIFFERENT CHAINS (A TO D).
 CC -1- SURCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: STALK CELLS.
 CC -1- SIMILARITY: THE SEQUENCE OF RESIDUES 1-149 OF THE B CHAIN IS
 CC IDENTICAL WITH THAT OF THE C CHAIN.
 CC -1- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.

CC -1- SURCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: STALK CELLS.
 CC -1- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

DR EMBL: J01282; AAA33197.1; -;
 DR PIR: A03381; DLD01A.
 DE DictyDB: D002080; dscA.
 DE DictyDB: TFE030421; FAS8_C.
 DR Pfam: PF00754; F5_F8_type_C; 1.
 DR SMART: SM00231; FAS8C; 1.
 DR PROSITE: PS01285; FAS8C_1; 1.
 DE PROSITE, PS01286; FAS8C_2; FAS8C_NIG.
 KW Lectin; Cell adhesion.
 FT DOMAIN 2 151
 FT SITE 79 81
 FT SITE 81 151
 FT SITE 151 253
 SQ SEQUENCE 253 AA: 28239 MW: 812634100F1FE91E GRC64;
 CELL ATTACHMENT SITE.

Query Match 67.48; Score 29; DB 1; Length 253;
 Best Local Similarity 62.54; Pred. No. 23;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QFVASYKV 9
 I I I I I
 DB 84 QWVTSYKI 91

RESULT 14
 DISC_DICDI
 ID DISC_DICDI STANDARD; PRT; 253 AA.
 AC P02887;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Discoidin I, C chain and B chain.
 GN DSCB AND DSCC.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A. (C CHAIN AND B CHAIN 1-149).
 RX MEDLINE=82170475; PubMed=6279874;
 RA Poole S., Firtel R.A., Lamar E., Rowekamp W.;
 RT "Sequence and expression of the discoidin I gene family in
 RT Dictyostelium discoideum."
 RL J. Mol. Biol. 153:273-289(1981).
 RN [2]
 RP CELL ATTACHMENT SITE.
 RX MEDLINE=85074463; PubMed=6509552;
 RA Springer W.R., Cooper D.N.W., Barondes S.H.;
 RT "Discoidin I is implicated in cell substratum attachment and ordered
 RT cell migration of Dictyostelium discoideum and resembles
 RT fibronectin."
 RL Cell 39:557-564(1984).
 CC -1- FUNCTION: GALACTOSE- AND N-ACETYLGLACTOSAMINE-BINDING LECTIN. MAY
 CC PLAY A ROLE IN CELL-SUBSTRATUM ADHESION RATHER THAN IN CELL-CELL
 CC ADHESION. MAY BE NECESSARY FOR THE MAINTENANCE OF NORMAL ELONGATE
 CC MORPHOLOGY DURING AGGREGATION.
 CC -1- SUBUNIT: TETRAMER OF FOUR DIFFERENT CHAINS (A TO D).
 CC -1- SURCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: STALK CELLS.
 CC -1- SIMILARITY: THE SEQUENCE OF RESIDUES 1-149 OF THE B CHAIN IS
 CC IDENTICAL WITH THAT OF THE C CHAIN.
 CC -1- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/anonymous/>
 CC or send an email to license@isb-sib.ch)

CC -----
 DR EMBL: J01284; AAA33190.1;
 DR EMBL: J01283; AAA33190.1;
 DR PIR: A03382; DLD01C.
 DR PIR: B03382; B03382.
 DR SWISS-2DPAGE: F02887; DICTY.
 DR DictyDb: DD02001; dscC.
 DR InterPro: IPR000421; F02887.
 DR SMART: SM00231; F02887; 1.
 DR PROSITE: PS01285; F02887; 1.
 DR PROSITE: PS01286; F02887; 1.
 DR Lectin: Cell adhesion.
 DR DOMAIN 2 151 F5/R TYPE C
 FT SITE 79 81 CELL ATTACHMENT SITE.
 FT SEQUENCE 253 AA; 2899; MW: 8756A94811F85444 CRC64,
 CC -----

Query Match 67.4%; Score 29; DB 1; Length 253;

Best Local Similarity 62.5%; Pred. No. 23;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0.

QY 2 QVASYKV ?

DB 84 QWTSYKI 91

RESULT 15

ID CYST_ECOLI STANDARD: PPT; 277 AA.
 AC P16701;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-MAP-2002 (Rel. 41, Last annotation update)
 DE Sulfate transport system permease protein cyst.
 GN CYST GP CYST GP R2424
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia
 OC NCBI_TaxID=562;
 PN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 PX MEDLINE=90254334; PubMed=2189958,
 RA Sirko A., Hryniewicz M M., Hulanicka O M., Ruck A.;
 RT "Sulfate and thiosulfate transport in Escherichia coli K-12;
 RT nucleotide sequence and expression of the cystWAM gene cluster.";
 RL J. Bacteriol. 172:3351-3357(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 PX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 PA Riley M., Collado-Vides J., Glasner J.D., Rode C.F., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H A., Goeden M A., Rose D J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1232-1244(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 PX MEDLINE=9749980; PubMed=9205837;
 RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
 RA Mizobuchi K., Muri H., Nakade S., Nakamura Y., Nishimoto H.,
 RA Oshima T., Oyama S., Saito N., Sano Y., Sivasubram S.,

EA Lejani S., Takahashi S., Fukuda J., Takemoto K., Uehara K., Wada C.,
 RA Yamada S., Horiuchi T.;
 FT "Construction of a contiguous 874-bb sequence of the Escherichia coli
 FT 812 genome corresponding to 50.0-58.8 min on the linkage map and
 FT analysis of its sequence features.";
 ML DNA Res. 4:91-113(1997)
 CC ! FUNCTION: PART OF THE RINGING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
 CC FOR SULFATE AND THIOSULFATE. PROBABLY RESPONSIBLE FOR THE
 CC TRANSPORTATION OF THE SULFATE ACROSS THE MEMBRANE.
 CC ! SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (potential).
 CC ! SIMILARITY: WITH INTEGRAL MEMBRANE COMPONENTS OF OTHER RINGING-
 CC PROTEIN-DEPENDENT TRANSPORT SYSTEMS. BELONGS TO THE CYSTW
 CC SUBFAMILY.
 CC -----

CC This SWISS PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/anonymous/>
 CC or send an email to license@isb-sib.ch)

CC -----

DR EMBL: M2101; AAA33637.1;
 DR EMBL: A6000330; AAC75477.1;
 DR EMBL: D56871; BAA16298.1;
 DR EMBL: D56872; BAA16307.1;
 DR PIR: A5402; APEPST.
 DR PIR: B35403; B35403.
 DR EMBL: E31797; cystW
 DR InterPro: IPR000515; RPD transp;
 DR Pfam: PF00528; RPD transp; 1.
 DR PROSITE: PS00402; RPD_TRANSP_INN_MEMBR; 1.
 KW Inner membrane; Transmembrane; Sulfate transport; Transport;
 KW Complete proteome;
 FT TRANSMEM 17 37 POTENTIAL.
 FT TRANSMEM 64 84 POTENTIAL.
 FT TRANSMEM 99 119 POTENTIAL.
 FT TRANSMEM 136 156 POTENTIAL.
 FT TRANSMEM 185 205 POTENTIAL.
 FT TRANSMEM 215 235 POTENTIAL.
 FT TRANSMEM 243 263 POTENTIAL.
 SQ SEQUENCE 277 AA; 3029; MW: 1392821B0DE4459 CRC64;

Query Match 67.4%; Score 29; DB 1; Length 277;

Best Local Similarity 75.0%; Pred. No. 26;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 QVASYKV 9

DB 55 QVAAVAVKV 62

Search completed: September 5, 2002, 15:31:43

Job time: 478 sec

GenCore version 4.5
Copyright (c) 1993-2000 Compaq, Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:31:52, Search time 12.673 seconds
(without alignments)
12.673 Million cell updates/sec

Title: US-09-744-804-40
Perfect score: 43
Sequence: 1 VQFVASYKV 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 56222 seqs, 17299129 residues
Total number of hits satisfying chosen parameters: 56222

Minimum DB seq length: 0
Maximum DP seq length: 20000000

Post-processing: Minimum Match 9%
Maximum Match 100%
Listing first 45 summaries

Database: SPTRMBL_19:

- 1: sp.archaea:*
- 2: sp.bacteria:*
- 3: sp.fungi:*
- 4: sp.human:*
- 5: sp.invertebrate:*
- 6: sp.mammal:*
- 7: sp.mhc:*
- 8: sp.organelle:*
- 9: sp.phage:*
- 10: sp.plant:*
- 11: sp.protein:*
- 12: sp.virus:*
- 13: sp.vertebrate:*
- 14: sp.unclassified:*
- 15: sp.yeast:*
- 16: sp.bacteriophage:*
- 17: sp.sparcheap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	EB ID	Description
1	39	90.7	426	11 Q9WTS3	Q9WTS3 mus musculus
2	39	90.7	463	11 Q9R1X9	Q9R1X9 mus musculus
3	36	83.7	480	4 O43854	O43854 homo sapien
4	36	83.7	480	11 Q5474	Q5474 mus musculus
5	32	74.4	330	5 Q963N3	Q963N3 drosophila
6	32	74.4	493	5 Q9VNA1	Q9VNA1 drosophila
7	32	74.4	610	5 Q9VCS1	Q9VCS1 drosophila
8	31	72.1	151	12 Q91922	Q91922 culicx nigri
9	31	72.1	213	16 Q9KTE8	Q9KTE8 vibrio chol
10	31	72.1	237	3 Q94290	Q94290 schistosom
11	31	72.1	308	16 Q9PFS9	Q9PFS9 xylella fas
12	31	72.1	313	4 Q9AW19	Q9AW19 homo sapien
13	31	72.1	344	2 Q66257	Q66257 actinobacil
14	31	72.1	1296	11 Q91216	Q91216 mus musculus
15	31	72.1	1226	17 Q58836	Q58836 methanococ
16	30	64.4	134	5 Q9VQ29	Q9VQ29 drosophila

17	30	69.8	159	10 Q94R10	Q94R10 vitis vinif
18	30	69.8	193	9 Q9MCE8	Q9MCE8 bacterioph
19	30	69.8	260	14 Q9P254	Q9P254 bacillus ba
20	30	69.8	290	10 Q92R10	Q92R10 arabidopsis
21	30	69.8	296	10 Q9Q060	Q9Q060 arabidopsis
22	30	69.8	358	10 Q7H381	Q7H381 lycopersico
23	30	69.8	363	6 Q77718	Q77718 equus cabal
24	30	69.8	374	16 Q9P626	Q9P626 xylella fas
25	30	69.8	386	17 Q9V169	Q9V169 pyrococcus
26	30	69.8	406	3 Q9H757	Q9H757 salmone
27	30	69.8	476	4 Q9U124	Q9U124 homo sapien
28	30	69.8	476	4 Q9U382	Q9U382 homo sapien
29	30	69.8	485	2 Q93171	Q93171 escherichia
30	30	69.8	529	5 Q18748	Q18748 caenorhabdi
31	30	69.8	576	5 P91296	P91296 caenorhabdi
32	30	69.8	672	5 Q92164	Q92164 caenorhabdi
33	30	69.8	967	5 Q97305	Q97305 plasmodium
34	30	69.8	1216	3 Q9V901	Q9V901 emericella
35	30	69.8	2162	5 Q9W2V6	Q9W2V6 drosophila
36	29	67.4	57	6 Q9N163	Q9N163 papio banad
37	29	67.4	106	12 Q98499	Q98499 paramecium
38	29	67.4	130	10 Q9FRM0	Q9FRM0 arabidopsis
39	29	67.4	131	8 Q9TA16	Q9TA16 caeteria r
40	29	67.4	214	8 Q9M111	Q9M111 paraconimus
41	29	67.4	222	12 Q99735	Q99735 sigma virus
42	29	67.4	223	12 Q64870	Q64870 porcine ade
43	29	67.4	229	2 Q57712	Q57712 staphylococ
44	29	67.4	274	5 Q965M8	Q965M8 caenorhabdi
45	29	67.4	276	2 Q40833	Q40833 rhodobacter

ALIGNMENTS

RESULT 1

ID Q9WTS3 PRELIMINARY: PRI: 426 AA.
AC Q9WTS3;
DI 01 NOV 1999 (TRMBLrel, 12, Created)
DI 01 NOV 1999 (TRMBLrel, 12, Last sequence update)
DI 01 DEC 2001 (TRMBLrel, 19, Last annotation update)
DE MILK FAT GLOBULE GLYCOPROTEIN MF5-E8 S (SIMILAR TO MILK FAT GLOBULE-EGF FACTOR 8 PROTEIN).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
FR STEATIN-HAIRLESS, TISSUE-MAMMARY GLAND,
RX MEDLINE-99120994; PubMed 9920772;
RA Oshima K., Aoki N., Negi M., Kitajima K., Matsuda T.,
RT "Lactation-dependent expression of an mRNA splice variant with an exon
for a multiply O-glycosylated domain of mouse milk fat globule
glycoprotein MF5-E8.";
RL Biochem. Biophys. Res. Commun. 254:522-528(1999).
RN [2]
RP SEQUENCE FROM N.A.
FR Strausberg R.;
RA Submitted (JEP 2001) to the EMBL/GenBank/DDBJ databases.
CL Submitted (JEP 2001) to the EMBL/GenBank/DDBJ databases.
CC - SIMILARITY: CONTAINS 2 P5/P8 TYPE C DOMAINS.
DR EMBL; AB025280; BAA76386.1; -;
DR PMEL; BC003004; AA030004.1; -;
DR EMBL; B2004892; AA034892.1; -;
DR HSPF; F00740; IEDM;
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001438; EGF_11.
DR InterPro; IPR000421; FAS5_C.
DR InterPro; IPR01072; HLB_dim.
DR Pfam; PF000308; EGF; 2.
DR Pfam; PF00754; F5_F8_type_C; 2.
DR PRINTS; PR00010; EGFBL00D.
DR SMART; SM00181; EGF; 2.

Best Local Similarity 77.8%; Pred. No. 15;
Matches 7; Conservative 1; Mismatches 1; Indels 3; Gaps 0.

QY 1 VOFVASYKV 9
11111111

Db 402 VOFVGSYKL 410

RESULT 4
ID Q35474 PRELIMINARY; PRT; 480 AA.
AC Q35474; Q35475;
DI 01-JAN-1998 (TREMREL. 05, Created)
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)
DE 01-DEC-2001 (TREMREL. 19, Last annotation update)
DE INTEGRIN BINDING REPEAT (F5/F8 TYPE C)
CN EDL3 OF DEL.
OS Mus musculus (Mouse).
CC Pakaryofa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND CHARACTERIZATION.
RC TISSUE=EMBRYO;
RX MEDLINE=98083109; PubMed 9420328;
RA Hidaï C., Zupancic T., Penta K., Mikhail A., Kawana M.,
RA Quertemous E.E., Aoka Y., Fukagawa M., Matsui Y., Platika D.,
RA Auerbach R., Hogan B.L.M., Snodgrass R., Quertemous T.;
FT "Cloning and characterization of developmental endothelial locus-1: an
RI embryonic endothelial cell protein that binds the alphavbeta3 integrin
RI receptor.";
PL Genes Dev. 12:21-33(1998).
CC -1- FUNCTION: PROMOTES ADHESION OF ENDOTHELIAL CELLS THROUGH
CC INTERACTION WITH THE ALPHA-V BETA3 INTEGRIN RECEPTOR. INHIBITS
CC FORMATION OF VASCULAR-LIKE STRUCTURES. MAY BE INVOLVED IN
CC REGULATION OF VASCULAR MORPHOGENESIS OR REMODELING IN EMBRYONIC
CC DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS: LONG ISOFORM (SHOWN HERE) AND
CC SHORT ISOFORM; ARE PROTECTED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ANGIOBLASTS AND EARLY ENDOTHELIAL
CC CELLS. BY EMBRYONIC DAY 13.5, ALSO EXPRESSED IN A RESTRICTED GROUP
CC OF NON-ENDOTHELIAL CELLS INCLUDING CHONDROCYTES AND RETINAL
CC NEURONS.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE EMBRYO FROM DAY 7, AFTER DAY
CC 15.5, EXPRESSION DECREASES AND DISAPPEARS COMPLETELY BY THE TIME
CC OF BIRTH.
CC -1- SIMILARITY: CONTAINS 3 EGF LIKE DOMAINS.
CC FMBL: AF031524; AAB86585,1; 2.
DR FMBL: AF031525; AAB86586,1; 2.
DR BEST: J03753, 1777.
DR MGI: MGI:129025; Ed113.
DR InterPro: IPR000561, ASX_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR01881; EGF_Ca.
DR InterPro: IPR000421; FAS5_C.
DR Pfam: PF00008; EGF_3.
DR Pfam: PF00754; F5_F8_type_C; 2.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00001; EGF-like; 2.
DR SMART: SM00231; FAS5C; 2.
DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
DR PROSITE: PS00322; EGF_1; UNIPROT_3.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS01285; FAS5C_1; 2.
DR PROSITE: PS01286; FAS5C_2; 2.
DR PROSITE: PS01286; FAS5C_2; 2.
KW EGF-like domain; Alternative splicing, signal, Developmental protein,
KW Cell adhesion; Repeat; Vascularization.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 480 INTEGRIN-BINDING PROTEIN DEL1.

FT DOMAIN 26 59 EGF-LIKE 1.
FT DOMAIN 79 116 EGF-LIKE 2.
FT DOMAIN 123 154 EGF-LIKE 3.
FT DOMAIN 161 411 F5/F8 TYPE C 1.
FT DOMAIN 322 473 F5/F8 TYPE C 2.
FT SITE 96 98 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 26 37 BY SIMILARITY.
FT DISULFID 31 48 BY SIMILARITY.
FT DISULFID 50 59 BY SIMILARITY.
FT DISULFID 78 89 BY SIMILARITY.
FT DISULFID 84 105 BY SIMILARITY.
FT DISULFID 107 116 BY SIMILARITY.
FT DISULFID 158 314 BY SIMILARITY.
FT DISULFID 301 405 BY SIMILARITY.
FT DISULFID 319 476 BY SIMILARITY.
FT VARSPFC 218 221 TRM 2 > VFWG (IN SHORT ISOPFORM).
FT VARSPFC 222 490 MISSING (IN SHORT ISOPFORM).
SQ SEQUENCE 480 AA, 53740 MW, 42693RTU, 2617145 GRC64;
Query Match 83.7%; Score 36; DB 11; Length 480;
Best Local Similarity 77.8%; Pred. No. 15;
Matches 7; Conservative 1; Mismatches 1; Indels 3; Gaps 0;
QY 1 VOFVASYKV 9
11111111
Db 402 VOFVGSYKL 410
RESULT 5
ID Q960N4 PRELIMINARY; PRT; 330 AA.
AC Q960N4;
DI 01-DEC-2001 (TREMREL. 19, Created)
DT 01-DEC-2001 (TREMREL. 19, Last sequence update)
DE 01-DEC-2001 (TREMREL. 19, Last annotation update)
DE ID43270P.
GN G11330.
OS Drosophila melanogaster (Fruit fly)
CC Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; insecta;
CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
CC Phyllophaga; Tephritidae; Drosophilidae;
CX NCBI_TaxID:7227;
RN [1]
RP SEQUENCE FROM N.A.
RC SEAIN.V. CN BW SP;
RA Stapleton M., Brokstein P., Hong L., Abmayani A., Carlson J.,
FA Chang H., Chaves C., Borsett V., Parian D., Frisc E., George P.,
FA Gonzalez M., Garfin H., Li P., Liao G., Miranda A., Mundall C.J.,
RA Nuno J., Paclob J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;
CC Submitted (2002-01-11) to EMBL, GenBank, DDBJ databases.
DR EMBL: AY051966; AAK93390,1; 2.
SQ SEQUENCE 330 AA, 37738 MW, 54951RTU, 4719292 GRC64;
Query Match 74.4%; Score 32; DB 5; Length 330;
Best Local Similarity 85.7%; Pred. No. 7;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 QFVASYK 8
11111111
Db 210 QFVASYK 216
RESULT 6
ID Q9VMA1 PRELIMINARY; PRT; 483 AA.
AC Q9VMA1;
DI 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
DE G51330 PROTEIN (CORTEX).

GN CG1330 OR CONT.

OS Drosophila melanogaster (Fruit fly)

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyceta; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10721172;

RA Adams M.F., Collier S.E., Holt P.A., Evans C.A., Sawyer J.D.,

RA Anandadas P.G., Scherter S.E., Li P.W., Hoskins P.A., Galle R.F.,

RA George P.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.P., Yandell M.D., Zhang Q., Chen L.X.,

RA Branton P.C., Rogers Y.H.C., Blazer J.P.G., Champ M., Pfeiffer B.D.,

RA Wan K.H., Boyjean C., Raxter E.G., Holt C., Nelson M., Miklos G.L.G.,

RA Abril J.F., Abayanti A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballou R.M., Basu A., Baxendale J., Bayraktarolu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Bertram R.P., Bhadani D., Bolshakov S.,

RA Borokova D., Borhan M.P., Bouck J., Brockstein P., Brothier P.,

RA Hurlis K.C., Busam D.A., Butler H., Gadieu E., Center A., Chandra I.,

RA Cherry J.M., Galloway S., Dahlke C., Bayenport L.R., Davies P.,

RA de Pablos R., Delcher A., Deng C., Mays A.D., Dew T., Dietz S.M.,

RA Durbin K.J., Evansellista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Glodok A., Gough F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.P., Houch J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Hengam C.,

RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennish J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Mout S.M., May M., Murphy R., Murphy L., Murry D.M., Nelson D.L.,

RA Merkulov G., Milshina N.V., Mobarry R., Murphy L., Murry D.M., Nelson D.L.,

RA Nelson D.K., Nelson K.A., Nixon K., Nusskern D.P., Pauley J.M.,

RA Palatino M., Pittman C.S., Pan S., Pollard T., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong P., Sun E.,

RA Svirkas R., Teeter C., Turner P., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.

FT "the genome sequence of Drosophila melanogaster";

RL Science 287:2185-2195(2000)

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21150807;

RA Chu T., Henrich G., Haegeli V., Strickland S.

FT "Context, a Drosophila gene required to complete oocyte meiosis, is a

PL member of the 24-141-152(fairy) protein family";

PL Genes 24:141-152(2001).

CC -1- SIMILARITY: CONTAINS 2 WB REPEATS (TPP-ASP DOMAINS)

DR EMBL: A630363.1; AAF52431.1;

DR EMBL: AY033478; AAK54464.1;

DR FlyBase: FBgn0003571; cont

DR InterPro: IPR001580; W440

DR Pfam: PF00400; W440; 2

DR SMART: SM00420; W440; 4

DR PROSITE: P550042; WD_PPFATS-2; 2

DR PROSITE: P550294; WD_PPFATS_PEGION; 1;

KW repeat; WD repeat.

SO SEQUENCE 483 AA; 55574 MW; 607100190606735 CYSK4;

Query Match 74.4%, Score 32, PR 5, length 483,

Best local Similarity 85.7%; pred. No. 1, loc 02;

Matches 6, conservative; 1 Mismatch; 0 Indels; 0 Gaps;

QY 2 QFVASYK 8

DB 363 QFVASYR 369

|||||

RESULT 7

Q9VC31

ID Q9VC31 PPFITINAPY; PPT 610 AA.

AC Q9VC31

DT 01-MAY-2000 (FEBR01.13, Created)

DT 01-MAY-2000 (FEBR01.13, Last sequence update)

DT 01-SEP-2001 (FEBR01.13, Last annotation update)

DE CG13638 PROTEIN.

GN CG13638

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyceta; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10721172;

RA Adams M.F., Collier S.E., Holt P.A., Evans C.A., Sawyer J.D.,

RA Anandadas P.G., Scherter S.E., Li P.W., Hoskins P.A., Galle R.F.,

RA George P.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.P., Yandell M.D., Zhang Q., Chen L.X.,

RA Branton P.C., Rogers Y.H.C., Blazer J.P.G., Champ M., Pfeiffer B.D.,

RA Wan K.H., Boyjean C., Raxter E.G., Holt C., Nelson M., Miklos G.L.G.,

RA Abril J.F., Abayanti A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballou R.M., Basu A., Baxendale J., Bayraktarolu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Bertram R.P., Bhadani D., Bolshakov S.,

RA Borokova D., Borhan M.P., Bouck J., Brockstein P., Brothier P.,

RA Hurlis K.C., Busam D.A., Butler H., Gadieu E., Center A., Chandra I.,

RA Cherry J.M., Galloway S., Dahlke C., Bayenport L.R., Davies P.,

RA de Pablos R., Delcher A., Deng C., Mays A.D., Dew T., Dietz S.M.,

RA Durbin K.J., Evansellista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Glodok A., Gough F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.P., Houch J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Hengam C.,

RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennish J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Mout S.M., May M., Murphy R., Murphy L., Murry D.M., Nelson D.L.,

RA Merkulov G., Milshina N.V., Mobarry R., Murphy L., Murry D.M., Nelson D.L.,

RA Nelson D.K., Nelson K.A., Nixon K., Nusskern D.P., Pauley J.M.,

RA Palatino M., Pittman C.S., Pan S., Pollard T., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong P., Sun E.,

RA Svirkas R., Teeter C., Turner P., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.

FT "the genome sequence of Drosophila melanogaster";

RL Science 287:2185-2195(2000)

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21150807;

RA Chu T., Henrich G., Haegeli V., Strickland S.

FT "Context, a Drosophila gene required to complete oocyte meiosis, is a

PL member of the 24-141-152(fairy) protein family";

PL Genes 24:141-152(2001).

CC -1- SIMILARITY: TO PAS PROTEINS. BELONGS TO THE PAS SUBFAMILY.

DR EMBL: A630363.1; AAF52431.1;

DR EMBL: AY033478; AAK54464.1;

DR FlyBase: FBgn0003571; Rab.

DR InterPro: IPR001580; Pas_rnsfing

DR Pfam: PF00071; pas; 1;

DR PRINTS: PB00449; RASTRNSFWNG.

DR SMART: SM00175; RAP; 1;

KW GTP-binding; Lipoprotein.

SO SEQUENCE 610 AA; 64876 MW; 148800488420774 CYSK4;

Query Match 74.4%; Score 32; DB 5; Length 610;
 Best Local Similarity 66.7%; Pred. No. 41;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0.

QY 1 VQVASYKV 9
 : |||:|
 DB 400 LDFVATYKV 408

RESULT 8
 Q919J2 PRELIMINARY: PRT; 121 AA.
 ID Q919J2
 AC Q919J2
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CUN083 HYPOTHETICAL PROTEIN.
 GN CUN083
 OS Culex nigripalpus baculovirus.
 OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae.
 OX NCBI_TaxID=130556;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FLORIDA1997;
 EX MEDLINE=21486685; PubMed=11602755;
 RA McGee G.L., Tulman E.R., Ia Z., Ballinsky G.A., Moser B.A.,
 RA Becnel J.J., Rock D.L., Kutish G.F.;
 RT "Genome Sequence of a Baculovirus Pathogenic for Culex nigripalpus."
 RL J. Virol. 75-11157-11165 (2001)
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FLORIDA1997;
 RA Afonso C.L., Tulman E.R., Ia Z., Ballinsky G.A., Moser B.A.,
 RA Becnel J.J., Rock D.L., Kutish G.F.;
 RL Submitted (Jul-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF403736; AAK94161.1;
 KW hypothetical protein.
 SQ SEQUENCE 121 AA; 13655 MW; 79971FD649E5B78F CRC64;

Query Match 72.1%; Score 31; DB 12; Length 121;
 Best Local Similarity 66.7%; Pred. No. 41;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0.

QY 1 VQVASYKV 9
 : |||:|
 DB 105 VEFALYKV 113

RESULT 9
 Q9KTE8 PRELIMINARY: PRT; 213 AA.
 ID Q9KTE8
 AC Q9KTE8
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE RARE LIPOPROTEIN B.
 GN VC0954.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCBI_TaxID=656;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EL TOR N16951 / SKETTYPE 01;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg P.F., Fison T.A., Nelson W.C., Clayton P.A., Gwin M.L.,
 RA Dodson P.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.P., Nelson K.E., Peck T.D., Tettelin H., Pickardson D.,
 RA Ermolaeva M.D., Vamathevan J., Ross S., Qin H., Fraaije F., Sellers P.,
 RA McDonald L., Mierbach T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., McKelanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio

RI cholerae.";
 RG Nature 406:477-484 (2000).
 DR DMBL: A003417; AAF94116.1;
 DR 1168; VC0954;
 KW lipoprotein; Complete proteome.
 SQ SEQUENCE 213 AA; 24121 MW; 26C95F2890269922 CRC64;

Query Match 72.1%; Score 31; DB 16; Length 213;
 Best Local Similarity 85.7%; Pred. No. 75;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 FVASYKV 9
 : |||||:
 DB 113 FVASYKV 119

RESULT 10
 G94290 PRELIMINARY: PRT; 217 AA.
 ID G94290
 AC G94290
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PUTATIVE RNA BINDING PROTEIN.
 GN SPBC887.05C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Sclerogasteromycetes;
 OC Schizosaccharomycotales; Schizosaccharomycetaceae;
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN 972H-;
 RA Lyne M., Rajandream M.A., Barrell B.G., Lelaure V., Galibert F.;
 EL Submitted (Nov 1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL033388; CAA21890.1;
 DR HSPD: P11940; 1CWI.
 DR InterPro: IPR000504; RRM.
 DR Pfam: PF00076; rim; 1.
 DR SMART: SM00460; RRM; 1.
 DR PROSITE: PS0102; RRM; 1.
 LR PROSITE: PS0030; RRM_RNP_1; 1.
 SQ SEQUENCE 217 AA; 25310 MW; 4F428FA4ABAAC8F CRC64;

Query Match 72.1%; Score 31; DB 4; Length 217;
 Best Local Similarity 77.8%; Pred. No. 76;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 VQVASYKV 9
 : |||||:
 DB 104 VQVASYKV 112

RESULT 11
 Q9PFS4 PRELIMINARY: PRT; 308 AA.
 ID Q9PFS4
 AC Q9PFS4
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN XF0584.
 GN XF0584.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN 945C;
 RX MEDLINE=20365717; PubMed 10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,

RA Alvares P., Alves I.M.C., Araujo T.F., Rala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi F.D., Bordin S., Rowe J.M., Rlones M.P.S.,
 RA Buato M.P.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
 RA Coutinho N.B., Colombo C., Costa F.F., Costa M.C.P., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto F., Iacono G., El-Lorri H.,
 RA Faciniani A.P., Ferreira I.S., Ferreira V.C.A., Ferro T.A.,
 RA Fraja J.S., Franca S.C., Franco M.C., Frohme M., Furian L.P.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Inunkeira M.L., Kemper E.L., Kitajima T.P.,
 RA Krieger J.E., Kuraue E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.F., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.L., Martins E.M.F., Matsukuma A.Y.,
 RA Meck C.F.M., Miranca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A., Jr., Nóbrega F.G., Nunes D.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto R.R., Pereira G.A.G., Pereira H.A., Jr., Pesquero J.B.,
 RA Quadrio F.R., Roberto F.G., Rodrigues V., de Faria A.T.M.,
 RA da Rosa V.E., Jr., de Sa R.G., Sartelli F.V., Sawasaki H.F.,
 RA da Silva A.C.P., da Silva A.M., da Silva F.P., Silva W.A., Jr.,
 RA da Silveira J.F., Silvestri M.L., Siqueira W.J., da Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi G., Tsai S.M., Tsubaki M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Zelditch J., Zelditch J.C.,
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*";
 RL Nature 406:151-159(2000)
 DR EMBL: AE003904: AAP8334.1: -
 KW Hypothetical protein: Complete proteome.
 SQ SEQUENCE: 308 AA; 344 AA; 444 AA; MW: 4157066Da; pI: 4.74

Query Match 72.1% Score 31; DB 16; Length 308;
 Best Local Similarity 85.7% Pctd. No. 1.1e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 FVASYKV 9
 DB 97 FVASYRV 103
 |||||

RESULT 12
 ID Q9NM19 PPELIMINARY: PPT: 310 AA
 AC Q9NM19: -
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE CDNA F120821 FIS, CLONE AN50049 (FRAGMENT).
 OS Homo sapiens (Human).
 OC Eukaryota, Metazoa, Chordata, Cladista, Vertebrata, Euteleostomi,
 OC Mammalia, Eutheria, Primates, Catarrhini, Hominoidea, Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-ADIPOSE TISSUE;
 RA Tanigami A., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K.,
 RA Hirao M., Ohmori Y., Ota T., Suzuki Y., Ohayashi M., Nishi T.,
 RA Shibata T., Tanaka T., Nakamura Y., Isozaki T., Sugano S.,
 RT "NEDO human CDNA sequencing project";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK008238: BAA0190.1: -
 FT NON-TEP 310 310
 SQ SEQUENCE: 310 AA; 35495 MW; 5CD4332EE6517A9C CRC64;

Query Match 72.1% Score 31; DB 4; Length 310;
 Best Local Similarity 85.7% Pctd. No. 1.1e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QFVASYK 8
 DB 19 QFVASYQ 25
 |||||

RESULT 13
 ID Q96257 PPELIMINARY: PPT: 344 AA
 AC Q96257: -
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE TINF3RAL MEMBRANE ACETYLTRANSFERASE
 OS Actinobacillus actinomycetecomitans (Haemophilus
 actinomycetecomitans).
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Actinobacillus
 OX NCBI_TaxID=7114;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCYC9710;
 EX MEMBRANE 500-576A, PubMed 383500.
 RA Nakano Y., Yoshida Y., Yamashita Y., Koga T.,
 RT "A gene cluster for 6-deoxy-1-talar synthesis in Actinobacillus
 actinomycetecomitans";
 RI Biochim Biophys Acta 1442:409-414(1998).
 DP EMBL: AF010415: BAA2813.1: -
 DR InterPro: IPR002656; DUF33.
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF01757; DUF33_2.
 DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
 KW Transferase.
 SQ SEQUENCE: 344 AA; 40036 MW; 1D7R65376AE8B80 CRC64;

Query Match 72.1% Score 31; DB 2; Length 344;
 Best Local Similarity 62.5% Pctd. No. 1.3e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 QFVASYKV 9
 DB 203 QFIVTYKV 209
 |||||

RESULT 14
 ID Q91216 PPELIMINARY: PPT: 1206 AA
 AC Q91216: -
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE KIAA0903-LIKE PROTEIN.
 OS Mus musculus (Mouse).
 OC Eukaryota, Metazoa, Chordata, Cladista, Vertebrata, Euteleostomi,
 OC Mammalia, Eutheria, Rodentia, Soricimorphia, Muridae, Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RA Fuchs S., Resch K., Thiel C., Platzer M., Jockusch H.,
 RA Schmitt-John T.;
 RT "An inversion between mouse and human chromosomes 11/2p13-15 detected
 by high-resolution comparative mapping";
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF424597: AAL2406.1: -
 FT NON-TEP 1206 1206
 SQ SEQUENCE: 1206 AA; 136046 MW; 345A9A757CF0AC74 CRC64;

Query Match 72.1% Score 31; DB 11; Length 1206;
 Best Local Similarity 86.7% Pctd. No. 4e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QFVASYK 8
 DB 19 QFVASYQ 25
 |||||


```

RESULT 15
Q58836 PRELIMINARY: PRT; 1226 AA.
AC Q58836;
DT 01-JAN-1998 (TRENDEL, 05, Created)
DT 01-JAN-1998 (TRENDEL, 05, Last sequence update)
DT 01-OCT-2001 (TRENDEL, 18, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ1441.
GN MJ1441.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
GC Methanococcus.
GX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gorayne J.D.,
RA Overbeck K., Fierres F.F., Weinstock E.C., Merick S.H., Gluck A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Martiny-Baron T.F., Kelley J.M., Peterson J.P., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Faine B.P., Beredjisky M.,
RA Kleink H.P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschii."
RL Science 273:1058-1073(1996).
CC -/- SIMILARITY: STRONG TO P.DENITRIFICANS COBN AND M.JANNASCHII
CC M10907.
DR EMBL: U67585; AAB99452.1; -.
DR TIGR: MJ1441; -.
DR InterPro: IPR003672; CcBN/Mg_chelatase.
DR Pfam: PF02514; ccbn-Mg_chel; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1226 AA; 141327 MW; 8F7DCE4E36162AB9 CRC64;

```

```

Query Match 72.1%; Score 31; DB 17; Length 1226;
Best Local Similarity 71.4%; Pred. No. 5002;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 QFVASYK 8
Db 510 QYIASYK 516

```

```

Search completed: September 5, 2002, 15:31:07
Job time: 472 sec

```

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes that proper record-keeping is essential for transparency and accountability, particularly in financial matters. The text outlines various methods for organizing and storing data, including digital databases and physical filing systems. It also mentions the need for regular audits and reviews to ensure the integrity of the information.

2. The second section focuses on the role of communication in achieving organizational goals. It highlights the importance of clear and concise communication channels, both internally and externally. The text suggests implementing regular meetings and reports to keep all stakeholders informed and engaged. It also discusses the benefits of open communication, such as improved collaboration and faster problem-solving.

3. The third part of the document addresses the challenges of managing a large and diverse team. It provides strategies for effective delegation, ensuring that tasks are assigned to the right people with the necessary skills and resources. The text also covers the importance of providing ongoing training and development opportunities to keep the team motivated and up-to-date. Additionally, it discusses the need for a strong leadership structure and the role of the manager in guiding the team towards success.

4. The final section discusses the importance of maintaining a positive and productive work environment. It suggests implementing policies that promote work-life balance and employee well-being. The text also emphasizes the need for a culture of respect and inclusivity, where all team members feel valued and heard. It concludes by stating that a positive work environment is crucial for attracting and retaining top talent, ultimately leading to the success of the organization.

GenCore version 4.5
Copyright (c) 1994-2000 Compugen Ltd.

OM protein - protein search, using sw mod-1

Run on: September 5, 2002, 15:27:36 : Search time: 58.86 seconds
(without alignments)
3.735 Million cell updates/sec

Title: US-09-744-804-41
Perfect score: 44
Sequence: 1 FL:AALSGA 9

Scoring table: RIGSMM62
Gapop 10 0 0 Gapext 0 5

Searched: 231629 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 9%
Maximum Match 100%
Listing first 45 summaries

- Database : Issued_Patents_AA:*
- 1: /seq2_6/ptdata/2/13a/5A_COMP pep.*
 - 2: /seq2_6/ptdata/2/13a/5B_COMP pep.*
 - 3: /seq2_6/ptdata/2/13a/6A_COMP pep.*
 - 4: /seq2_6/ptdata/2/13a/6B_COMP pep.*
 - 5: /seq2_6/ptdata/2/13a/1C10S_COMP pep.*
 - 6: /seq2_6/ptdata/2/13a/backfiles pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	100.0	387	US-08-162-402B-6	Sequence 6, Appl
2	44	100.0	405	US-08-162-402B-6	Sequence 9, Appl
3	37	84.1	463	US-08-162-402B-9	Sequence 9, Appl
4	36	81.8	355	US-08-890-719-11	Sequence 11, Appl
5	36	81.8	355	US-08-890-719-13	Sequence 13, Appl
6	32	72.7	361	US-08-415-751-36	Sequence 36, Appl
7	31	70.5	15	US-08-190-802A-1	Sequence 1, Appl
8	31	70.5	15	US-08-477-346-1	Sequence 1, Appl
9	31	70.5	15	US-08-473-089-1	Sequence 1, Appl
10	31	70.5	88	US-09-216-295-17	Sequence 17, Appl
11	31	70.5	470	US-08-691-814B-2	Sequence 2, Appl
12	31	70.5	644	US-08-336-708A-9	Sequence 9, Appl
13	31	70.5	1210	US-08-484-438-7	Sequence 7, Appl
14	31	70.5	1210	US-08-475-035-4	Sequence 4, Appl
15	30	68.2	259	US-09-216-295-5	Sequence 5, Appl
16	30	68.2	324	US-09-194-905-11	Sequence 11, Appl
17	30	68.2	475	US-09-243-989-2	Sequence 2, Appl
18	30	68.2	476	US-09-233-989-3	Sequence 3, Appl
19	30	68.2	476	US-09-233-989-6	Sequence 6, Appl
20	29	65.9	27	US-09-142-469-3	Sequence 3, Appl
21	29	65.9	46	US-08-691-814B-17	Sequence 17, Appl
22	29	65.9	51	US-08-173-102-8	Sequence 8, Appl
23	29	65.9	51	US-08-173-102-8	Sequence 8, Appl
24	29	65.9	86	US-08-421-144A-9	Sequence 9, Appl
25	29	65.9	104	US-08-466-033-83	Sequence 83, Appl
26	29	65.9	104	US-08-444-733-63	Sequence 84, Appl
27	29	65.9	104	US-08-464-134-83	Sequence 83, Appl

28	29	65.9	104	2	US-08-461-361-83	Sequence 83, Appl
29	29	65.9	104	2	US-08-482-910-83	Sequence 83, Appl
30	29	65.9	104	5	US-08-482-910-83	Sequence 83, Appl
31	29	65.9	118	4	US-08-627-376-17	Sequence 17, Appl
32	29	65.9	122	1	US-08-627-376-17	Sequence 17, Appl
33	29	65.9	122	1	US-08-627-376-17	Sequence 17, Appl
34	29	65.9	122	1	US-08-627-376-17	Sequence 17, Appl
35	29	65.9	122	1	US-08-627-376-17	Sequence 17, Appl
36	29	65.9	125	4	US-08-475-879-37	Sequence 37, Appl
37	29	65.9	129	2	US-08-475-879-37	Sequence 37, Appl
38	29	65.9	132	4	US-08-785-065-3	Sequence 3, Appl
39	29	65.9	132	4	US-08-785-065-3	Sequence 3, Appl
40	29	65.9	143	2	US-08-676-444-47	Sequence 47, Appl
41	29	65.9	143	2	US-08-676-444-47	Sequence 47, Appl
42	29	65.9	153	4	US-08-611-757-105	Sequence 105, Appl
43	29	65.9	153	4	US-08-611-757-105	Sequence 105, Appl
44	29	65.9	159	2	US-09-372-122A-46	Sequence 46, Appl
45	29	65.9	190	2	US-09-372-122A-46	Sequence 46, Appl
46	29	65.9	190	2	US-09-372-122A-46	Sequence 46, Appl
47	29	65.9	190	2	US-09-372-122A-46	Sequence 46, Appl
48	29	65.9	190	2	US-09-372-122A-46	Sequence 46, Appl
49	29	65.9	190	2	US-09-372-122A-46	Sequence 46, Appl
50	29	65.9	190	2	US-09-372-122A-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1
US-08-162-402B-6
Sequence 6, Application US/08162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERIANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LAPOCCA, DAVID J.
TITLE OF INVENTION: 45 KILALON HUMAN MILK FAT
TITLE OF INVENTION: 45 KILALON HUMAN MILK FAT
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fectio, Schuchman & Fogelowski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
ALIGNMENT NUMBER: 02, 06, 15, 40, 5B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/SEARCH NUMBER: P66 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213 622-7700
TELEFAX: 213 489 4210
TELEX:
INFORMATION FOR SEQ ID NO. 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 387 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-6

Query Match: 100.0%, Score 44, 387, 100.0%
Best Local Similarity: 100.0%, E-Val: No. 0.75,

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLAAALCGA 9
|||||
DB 5 RLAAALCGA 13

RESULT 2

US-08-162-402B-8
Sequence 8 Application US/08162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERIANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LAROCCA, DAVID J.
TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT
TITLE OF INVENTION: GLOBULE (HMEG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplawski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08162.402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: P66 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 465 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-162-402B-8

Query Match 100.0%; Score 44; DB 2; Length 465;

Best Local Similarity 100.0%; Pred. No. 0.9;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

QY 1 RLAAALCGA 9
|||||
DB 5 RLAAALCGA 13

RESULT 3

US-08-162-402B-9
Sequence 9 Application US/08162402B
Patent No. 5972337
GENERAL INFORMATION:
APPLICANT: CERIANI, ROBERTO L.
APPLICANT: PETERSON, JERRY A.
APPLICANT: LAROCCA, DAVID J.
TITLE OF INVENTION: 46 KDALTON HUMAN MILK FAT

TITLE OF INVENTION: GLOBULE (HMEG) ANTIGEN, FRAGMENTS & FUSION PROTEIN
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder & Poplawski
STREET: 444 South Flower St., 19th Floor
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08162.402B
FILING DATE: 03-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Amzel, Viviana
REGISTRATION NUMBER: 30,930
REFERENCE/DOCKET NUMBER: P66 38215
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
TELEX:

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 463 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-162-402B-9

Query Match 84.1%; Score 47; DB 2; Length 463;

Best Local Similarity 87.5%; Pred. No. 17;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLAAALCG 8
|:|||||
DB 5 RVLAAALCG 12

RESULT 4

US-08-890-719-11

Sequence 11, Application US/08890719A

Patent No. 6075125

GENERAL INFORMATION:

APPLICANT: Bacon, Larry D

APPLICANT: Hunt, Henry D

APPLICANT: Fulton, Janet

TITLE OF INVENTION: Production of Antisera Specific to Major

FILE REFERENCE: DKT 0064.96 - Larry D. Bacon et al.

CURRENT APPLICATION NUMBER: US/08/890.719A

CURRENT FILING DATE: 1997-07-09

EARLIER APPLICATION NUMBER: 60/021,685

EARLIER FILING DATE: 1996-07-10

NUMBER OF SEQ ID NOS: 39

SOFTWARE: Patent In Ver. 2.1

SEQ ID NO 11

LENGTH: 355

TYPE: PRT

ORGANISM: Gallus gallus

US-08-890-719-11

Query Match 81.8%; Score 36; DB 3; Length 355;

Best Local Similarity 87.5%; Pred. No. 20;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLAALCGA 9
|||||
DB 12 LLAACGA 19

RESULT 5
US-08-890-719-13
Sequence 13, Application US/08890719A
Patent No. 6075125
GENERAL INFORMATION:
APPLICANT: Bacon, Larry D
APPLICANT: Hunt, Henry D
APPLICANT: Fulton, Janet
TITLE OF INVENTION: Production of Antisera Specific to Major
FILE REFERENCE: Dkt 0064.96 - Larry D. Bacon et al.
CURRENT APPLICATION NUMBER: US/08/890,719A
CURRENT FILING DATE: 1997-07-09
EARLIER APPLICATION NUMBER: 607021,685
EARLIER FILING DATE: 1996-07-10
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 13
LENGTH: 355
TYPE: PRT
ORGANISM: Gallus gallus
US-08-890-719-13

Query Match 81.8%; Score 36; DB 3; Length 355;
Best Local Similarity 87.5%; Pred. No. 20;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLAALCGA 9
|||||
DB 12 LLAACGA 19

RESULT 6
US-08-415-751-36
Sequence 36, Application US/08415751
Patent No. 5643772
GENERAL INFORMATION:
APPLICANT: PETERSEN, CAROLYN
APPLICANT: LEECH, JAMES
APPLICANT: NELSON, RICHARD, C.
APPLICANT: GUT, JTRI
TITLE OF INVENTION: POLYPEPTIDES BINDING ANTIL-
TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA
TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID
TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND
TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND
TITLE OF INVENTION: DIAGNOSIS AND KIT
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: PHILLIPS, MOORE, LEMPLO & FINLEY
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: California
COUNTRY: United States of America
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage
COMPUTER: PC
OPERATING SYSTEM: IWS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/415,751
FILING DATE: 03-APR-1995
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/071,880
FILING DATE: June 1, 1993
APPLICATION NUMBER: 07/891,401
FILING DATE: May 29, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Hana Dolezalova
REGISTRATION NUMBER: 30,518
REFERENCE/AGENT NUMBER: 480,19-2 (HHD)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-1677
TELEFAX: (415) 324-1678
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE: Cryptosporidium parvum
FEATURE:
NAME/KEY: Positions coded by nonsense codons are
NAME/KEY: identified as Xaa.
US-08-415-751-36

Query Match 72.7%; Score 32; DB 1; Length 361;
Best Local Similarity 62.5%; Pred. No. 1,1e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KLAALCG 8
|:|:|
DB 44 KILAGICG 51

RESULT 7
US-08-190-802A-1
Sequence 1, Application 03/08190802A
Patent No. 5519003
GENERAL INFORMATION:
APPLICANT: Mochly-Rosen, Daria
APPLICANT: Ron, Dorit
TITLE OF INVENTION: WD-40 - Derived peptides and Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dellinger & Associates
STREET: P.O. Box 60850
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-0850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/190,802A
FILING DATE: 01-FEB-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/AGENT NUMBER: 8600 0139
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid

```

; Topology: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Peptide I
US-08-140-802A-1

Query Match 70.5%; Score 31; DB 1; Length 15;
Best Local Similarity 62.5%; Pred. No. 7.5;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLAAALCG 8
Db 6 KILVALCG 13

RESULT 8
US-08-477-346-1
; Sequence 1, Application US/08477346
; Patent No. 6262029
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,346
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,072
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/WORKET NUMBER: 2550-0025 20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Peptide I
US-08-477-446-1

Query Match 70.5%; Score 31; DB 4; Length 15;
Best Local Similarity 62.5%; Pred. No. 7.5;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLAAALCG 8
Db 6 KILVALCG 13

RESULT 8
US-08-477-346-1
; Sequence 1, Application US/08477346
; Patent No. 6262029
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,346
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,072
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/WORKET NUMBER: 2550-0025 20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Peptide I
US-08-477-446-1

```

```

Db 6 KILVALCG 13

RESULT 9
US-08-473-089-1
; Sequence 1, Application US/08473089
; Patent No. 6342368
; GENERAL INFORMATION:
; APPLICANT: Mochly-Rosen, Daria
; APPLICANT: Ron, Dorit
; TITLE OF INVENTION: WD-40 - Derived Peptides and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,089
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/WORKET NUMBER: 2550-0025 22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Peptide I
US-08-473-089-1

Query Match 70.5%; Score 31; DB 4; Length 15;
Best Local Similarity 62.5%; Pred. No. 7.5;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLAAALCG 8
Db 6 KILVALCG 13

RESULT 10
US-09-216-295-17
; Sequence 17, Application US/09216295
; Patent No. 6268328
; GENERAL INFORMATION:
; APPLICANT: Mitchinson, Colin
; APPLICANT: Wondt, Dan J.
; TITLE OF INVENTION: No. 6268328a1 Variant EGIII-Like Collulase Compositions
; FILE REFERENCE: GC555
; CURRENT APPLICATION NUMBER: US/09/216,295
; CURRENT FILING DATE: 1998-12-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSQ for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 88

```

TYPE: PRT
ORGANISM: Gliocladium roseum (3)
US-09-216-295-17

Query Match 70.5% Score 31; DB 4; Length 88;
Best local Similarity 100.0% Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LAALCG 8
DB 16 LAALCG 21

RESULT 11
US-08-691-814B 2
Sequence 2, Application US/08691814B
Patent No. 5981218
GENERAL INFORMATION:
APPLICANT: Rio, Marie-Christine
APPLICANT: Tomasette, Catherine
APPLICANT: Basset, Paul
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful
as Leukemia Markers and in Breast Cancer Diagnosis
NUMBER OF SEQUENCES: 124
CORRESPONDENCE ADDRESS:
ADDRESS: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Ave, NW, Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3934

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691-814B
FILING DATE: 31-JUL-1996
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/002,183
FILING DATE: 09-AUG-1995

ATTORNEY/AGENT INFORMATION:
NAME: Stoffe, Eric K.

REGISTRATION NUMBER: 46,682
REFERENCE/DOCKET NUMBER: 1383.0090001

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600

TELEFAX: 202-371-2643

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 470 amino acids

TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-691-814B-2

Query Match 70.5% Score 31; DB 2; Length 470;
Best local Similarity 76.0% Pred. No. 230-02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KLLAALCG 8
DB 15 KLLAALCG 22

RESULT 12
US-08-736-708A-9
Sequence 4, Application US/08736708A

Patent No. 5521295
GENERAL INFORMATION:
APPLICANT: Pacific, Robert E.
APPLICANT: Thomson, Arlen R.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: Hybrid Receptor Molecules
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Anden Inc.
STREET: 1840 behavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/736-708A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gleski, Nancy
REFERENCE/DOCKET NUMBER: A-241A
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 644 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-336-708A-9

Query Match 70.5% Score 31; DB 1; Length 644;
Best local Similarity 87.5% Pred. No. 31e-02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLAALCG 9
DB 14 LLAALCG 21

RESULT 13
US-08-484-438-7
Sequence 7, Application US/08484438
Patent No. 5811098
Patent No. 5811098 5780031

GENERAL INFORMATION:
APPLICANT: Pleszars, Gregory D.

APPLICANT: Culouscou, Jean-Michel

APPLICANT: Shoyab, Mohammed

APPLICANT: Siedail, Clay B.

APPLICANT: Hellstr m, Ingegerd

APPLICANT: Hellstr m, Karl F.

TITLE OF INVENTION: Hk4 HUMAN RECEPTOR TYROSINE KINASE

NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,438

1 FILING DATE: 07-JUN-1995
2 CLASSIFICATION: 530
3 PRIOR APPLICATION DATA:
4 APPLICATION NUMBER: 08/123,442
5 FILING DATE: 14-OCT-1994
6 APPLICATION NUMBER: US 08/150,704
7 FILING DATE: 10-NOV-1993
8 CLASSIFICATION: 530
9 PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: US 07/981,165
11 FILING DATE: 24-REV-1992
12 CLASSIFICATION: 530
13 ATTORNEY/AGENT INFORMATION:
14 NAME: Mistrock, S. Leslie
15 REGISTRATION NUMBER: 18,872
16 REFERENCE/DOCKET NUMBER: 5624-230
17 TELECOMMUNICATION INFORMATION:
18 TELEPHONE: (212) 790-9090
19 TELEFAX: (212) 869-8864/9741
20 TELEFAX: 66141 PENNIE
21 INFORMATION FOR SEQ ID NO. 7:
22 SEQUENCE CHARACTERISTICS:
23 LENGTH: 1216 amino acids
24 TYPE: amino acid
25 STRANDEDNESS: unknown
26 TOPOLOGY: unknown
27 MOLECULE TYPE: protein
28 US-08-484-438-7

Query Match 70.5%; Score 31; DB 2; Length 1210;
Best Local Similarity 87.5%; Pred No. 5.7e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLAALCGA 9
DB 14 LLAALCPA 21

RESULT 14
US-08-475-035-4
1 Sequence 4, Application US/08475035
2 Patent No. 598553
3 GENERAL INFORMATION:
4 APPLICANT: KING, C. R.
5 APPLICANT: KRAUS, MATTHIAS H.
6 APPLICANT: AARONSON, STUART A.
7 TITLE OF INVENTION: HUMAN GENE RELATED TO RUT DISTINCT FROM
8 TITLE OF INVENTION: EGF RECEPTOR GENE
9 NUMBER OF SEQUENCES: 4
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: NEEDLE & POSENBERG, P.C.
12 STREET: Suite 1200, 127 Peachtree Street
13 CITY: Atlanta
14 STATE: Georgia
15 COUNTRY: USA
16 ZIP: 30303
17 COMPUTER READABLE FORM:
18 MEDIUM TYPE: Floppy disk
19 COMPUTER: IBM PC compatible
20 OPERATING SYSTEM: PC-DOS/MS-DOS
21 SOFTWARE: Patent In Release #1.0, Version #1.30
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER: US/08/475,035
24 FILING DATE: 7 Jun 1995
25 CLASSIFICATION: 435
26 ATTORNEY/AGENT INFORMATION:
27 NAME: Perryman, David G.
28 REGISTRATION NUMBER: 33,438
29 REFERENCE/DOCKET NUMBER: 1414 656
30 TELECOMMUNICATION INFORMATION:
31 TELEPHONE: 404/688-0770
32 TELEFAX: 404/688-9880

1 INFORMATION FOR SEQ ID NO. 4:
2 SEQUENCE CHARACTERISTICS:
3 LENGTH: 1210 amino acids
4 TYPE: amino acid
5 TOPOLOGY: linear
6 MOLECULE TYPE: protein
7 US-08-475-035-4

Query Match 70.5%; Score 31; DB 2; Length 1210;
Best Local Similarity 87.5%; Pred No. 5.7e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLAALCGA 9
DB 14 LLAALCPA 21

RESULT 15
US-09-216-295-5
1 Sequence 5, Application US/09216295
2 Patent No. 626828
3 GENERAL INFORMATION:
4 APPLICANT: Mitchinson, Colin
5 APPLICANT: Wendt, Dan J.
6 TITLE OF INVENTION: No. 626828el Variat Cell-like Cellulase Compositions
7 FILE REFERENCE: GC555
8 CURRENT APPLICATION NUMBER: US/09/216,295
9 CURRENT FILING DATE: 1998-12-18
10 NUMBER OF SEQ ID NOS: 41
11 SOFTWARE: FastSeq for Windows Version 3.0
12 SEQ ID NO 5
13 LENGTH: 259
14 TYPE: PRT
15 ORGANISM: Aspergillus aculeatus
16 US-09-216-295-5

Query Match 68.2%; Score 30; DB 4; Length 259;
Best Local Similarity 87.5%; Pred No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLAALCGA 9
DB 6 LLAALAGA 13

Search completed: September 5, 2002, 15:27:37
Job time: 322 sec

GenCore version 4.5
Copyright (c) 1998 - 2000 Compugen Ltd.

QM protein - protein search, using sw med1

Run on: September 5, 2002, 15:36:32 ; Search time 158.43 seconds
(without alignments)
6.310 Million cell updates/sec

Title: US-09-744-804-41
Perfect score: 44
Sequence: 1 PUAALGGA 9

Scoring table: BIOSUM62
Gapop 10 0, Gapext 0 5

Searches: 747574 seqs, 11107456 residues
Total number of hits satisfying chosen parameters: 747574

Minimum OR seq length: 0
Maximum BK seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :			
A_Geneseq_032802.*			
1:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*	1	77.3
2:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*	2	77.3
3:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*	3	77.3
4:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*	4	77.3
5:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*	5	77.3
6:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*	6	77.3
7:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*	7	77.3
8:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*	8	77.3
9:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*	9	77.3
10:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*	10	77.3
11:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*	11	77.3
12:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*	12	77.3
13:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*	13	77.3
14:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*	14	77.3
15:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*	15	77.3
16:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*	16	77.3
17:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*	17	77.3
18:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*	18	77.3
19:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*	19	77.3
20:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*	20	77.3
21:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*	21	77.3
22:	/SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*	22	77.3

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	EA	ID	Description
1	44	100.0	9	21	AA198446	Lactadherin (HA-46)
2	44	100.0	487	16	AA197252	HMFG 46 kDa antigen
3	44	100.0	387	21	AA194453	Human lactadherin
4	37	84.1	426	21	AA194454	Mouse lactadherin
5	36	81.8	345	9	AA193149	Probe F10-encoded
6	36	81.8	355	21	AA194506	Chicken BF121 c1a
7	36	81.8	355	21	AA194507	Chicken BF121 c1a
8	35	79.5	365	22	ABG5914	Novel human diaph
9	34	77.3	160	22	AA040306	Propionibacterium
10	34	77.3	197	22	AA041965	Propionibacterium
11	34	77.3	466	22	AB170578	Drosophila melanog

12	34	77.3	1754	22	ABG19818	Novel human diaph
13	34	77.3	1759	22	AA032273	Human polyptide
14	33	75.0	80	22	AA185537	Human secreted pro
15	33	75.0	106	22	AA032273	Novel human diaph
16	33	75.0	218	22	AA194969	Myobacterium spec
17	33	75.0	218	22	AA194969	Myobacterium tube
18	33	75.0	218	22	AA194969	Myobacterium spec
19	33	75.0	218	22	AA194969	Myobacterium spec
20	33	75.0	218	22	AA194969	Myobacterium spec
21	33	75.0	218	22	AA194969	Myobacterium spec
22	33	75.0	218	22	AA194969	Myobacterium spec
23	33	75.0	218	22	AA194969	Myobacterium spec
24	33	75.0	218	22	AA194969	Myobacterium spec
25	33	75.0	218	22	AA194969	Myobacterium spec
26	33	75.0	218	22	AA194969	Myobacterium spec
27	33	75.0	218	22	AA194969	Myobacterium spec
28	33	75.0	218	22	AA194969	Myobacterium spec
29	33	75.0	218	22	AA194969	Myobacterium spec
30	33	75.0	218	22	AA194969	Myobacterium spec
31	33	75.0	218	22	AA194969	Myobacterium spec
32	33	75.0	218	22	AA194969	Myobacterium spec
33	33	75.0	218	22	AA194969	Myobacterium spec
34	33	75.0	218	22	AA194969	Myobacterium spec
35	33	75.0	218	22	AA194969	Myobacterium spec
36	33	75.0	218	22	AA194969	Myobacterium spec
37	33	75.0	218	22	AA194969	Myobacterium spec
38	33	75.0	218	22	AA194969	Myobacterium spec
39	33	75.0	218	22	AA194969	Myobacterium spec
40	33	75.0	218	22	AA194969	Myobacterium spec
41	33	75.0	218	22	AA194969	Myobacterium spec
42	33	75.0	218	22	AA194969	Myobacterium spec
43	33	75.0	218	22	AA194969	Myobacterium spec
44	33	75.0	218	22	AA194969	Myobacterium spec
45	33	75.0	218	22	AA194969	Myobacterium spec

ALIGNMENTS

RESULT 1	
AA182846	
ID	AA182846 standard; peptide: 9 AA.
XX	
AC	AA182846;
XX	
XX	19-JUN-2000 (first entry)
XX	
XX	Lactadherin (HA 46) peptide fragment (tumour associated antigen).
XX	
XX	Tumour associated antigen peptide; TAA; cancer; carcinoma;
XX	triple-antigen; prevention; cure; anti-tumour vaccine; metastases;
XX	breast; bladder; prostate; pancreas; ovary; thyroid; colon;
XX	stomach; carcinoma; MHC Class I; HLA-A2; human;
XX	Major Histocompatibility Complex; uroplakin;
XX	prostate specific antigen; prostate specific membrane antigen;
XX	prostate acid phosphatase, mucin, lactadherin;
XX	teratocarcinoma derived growth factor; PSA; PSMA; PAP; CRIPTO-1.
OS	Homo sapiens.
XX	
XX	WO200006723-A1.
XX	
XX	10-SEP-2000.
XX	
XX	29-JUL-1999, 3000-1100417.
XX	
XX	30-JUL-1998; 98IL-0125608.
XX	(YEDA) YEDA RES & DEV CO LTD.
XX	(RICE) BIO-TECHNOLOGY GEN CORP.
XX	Eisenbach L, Carmen L, Tirosh B, Bar-haim E, Paz A, Fridkin M,
XX	Filizer atlas C;

XX WPI: 2000-205463/18.
 XX Tumor associated antigen peptides, especially derived from uroplakin,
 PT useful as vaccines to prevent or cure cancers including breast,
 PT bladder, prostate, pancreas, ovary, thyroid, colon and stomach -
 XX
 XX Claim 17; Page 100; 11pp; English.
 XX
 CC Tumor associated antigen peptides (TAA) may be used for the
 CC treatment, prevention and cure of cancer or cancer metastases. The
 CC cancer may be breast, bladder, prostate, pancreas, ovary, thyroid,
 CC colon, stomach, head or neck cancer or a carcinoma. The tumor
 CC associated antigens are presentable to the immune system by HLA-A2
 CC molecules and are generally between 8 to 10 amino acids in length.
 CC The amino acids located at positions 2 and 9 of the tumor associated
 CC antigens are the anchor residues which participate in the binding to
 CC MHC class I molecules, more specifically HLA A2. More tumor
 CC associated antigens are described in GENESQ records AAY82806-Y82807.
 CC Those tumor associated antigens described in records AAY82806-Y82824
 CC and AAY82855-Y82869 are derived from uroplakin, such as uroplakin II,
 CC uroplakin Ia, uroplakin III and uroplakin Ib. Those described in
 CC records AAY82825-Y82829 are derived from prostate specific antigen
 CC (PSA). Those described in records AAY82830-Y82835 are derived from
 CC prostate specific membrane antigen (PSMA). Those described in
 CC records Y82836-AAY82839 are derived from prostate acid phosphatase
 CC (PAP). Those described in records AAY82840-Y82846 are derived
 CC from lactadherin (BA-46). Those described in records AAY82847-Y82854
 CC are derived from Mucin and those described in records AAY82871-Y82882
 CC are derived from luteal carcinoma derived growth factor (CRIPTO-1).
 XX
 SQ Sequence 9 AA;

Query Match 100.0%; Score 44; DB 21; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLAAALCGA 9
 DB 1 RLAAALCGA 9

RESULT 2
 AAR77252
 ID AAR77252 standard. Protein. 387 AA.
 AC AAR77252;
 XX

DE 21-NOV-1995 (first entry)

XX HMFG 46 kDa antigen.

XX HMFG; human milk fat globule; antigen; immunogen; vaccine; virucide;
 KW epithelium; tumor; breast cancer; monoclonal antibody; MAB

XX Homo sapiens.

XX W09515171-A.

XX 08-JUN-1995.

XX 05-DEC-1994; 94WO-051967.

XX 03-DEC-1993; 93US-0162402.

XX (CANC-) CANCER RES FUND CONTRA COSTA.

XX Coriani KL, Larocca RJ, Peterson JA;

XX WPI: 1995-215151/28.

XX N-PSDB; AA091198.

PT 46 kDa apparent molecular weight human milk fat globule antigen
 PT used in assays to determine the presence of a cancerous tumour of
 PT epithelial origin, and in a vaccine against neoplastic tumours
 XX
 XX claim 5, Page 40 47, 58pp; English.
 XX
 CC A complete cDNA sequence for the 46 kDa HMFG antigen, a major
 CC component of the apical surface of the normal breast epithelial
 CC cell, was obtained by PCR and RACE methods. cDNA clones can be
 CC used to prepare MABs for use in immunotherapy, immunohistochemistry,
 CC prognosis, diagnosis, imaging and therapy. Recombinant antigen can
 CC be expressed in prokaryotic or eukaryotic cells.

SQ Sequence 387 AA;

Query Match 100.0%; Score 44; DB 16; Length 387;
 Best Local Similarity 100.0%; Pred. No. 6.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLAAALCGA 9
 DB 5 RLAAALCGA 13

RESULT 3
 AAY94453
 ID AAY94453 standard. Protein. 387 AA.
 XX AAY94453;
 XX

XX 11-SEP-2000 (first entry)

XX Human lactadherin protein.

XX Human lactadherin; MGF-E8; anti-tumour; immune response;
 KW exosome; dendritic cell.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..23
 FT /label= Secretion_signal

FT Protein 24..387

FT /label= Lactadherin

FT Binding-site 46..48

FT /label= Inteinin_binding_site

XX EP1004664-A1.

XX 31-MAY-2000.

XX 24-NOV-1998; 98EP-0402925.

XX 24 NOV 1998; 98EP 0402925.

XX (INRM) INRM INST NAT SANTE & RECH MEDICALE.

XX (CUR1-) INSI CURIE.

XX WPI: 2000-352597/31.

XX N-PSDB; AAA27140.

XX Chimeric isolated (human) lactadherin polypeptide that functions as an
 PT adaptor of cross-priming to eliminate pathogenic antigens, e.g. in
 PT tumors -

XX Example 3; Page 12; 20pp; English.

XX Lactadherin protein was found in exosomes produced by dendritic cells.
 CC The protein is involved in the phagocytosis of particulate antigens by
 CC dendritic cells. Exosomes produced by dendritic cells exposed to
 CC tumour antigens induce potent immune responses. Lactadherin or variants
 CC of it may be used in the mediation of an immune response. Variants of

CC lactadherin may be used for inhibition and/or stimulation of the
 CC cross-priming of antigens and stimulation of the phagocytosis of
 CC antigens by dendritic cells. Compositions derived from lactadherin can
 CC also be used to monitor an immune response, more specifically a CTL
 CC (cytotoxic T lymphocyte) response and also to produce CTLs specific
 CC for a selected antigen. The present sequence is the human
 CC lactadherin protein.

XX Sequence 387 AA.

Query Match 100.0%, Score 44; DB 21; Length 387.
 Best Local Similarity 100.0%, Pred. No. 6.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLIAALCGA 9
 Db 5 RLIAALCGA 13

RESULT 4

AA94454
 ID AAY94454 standard; Protein; 426 AA.

XX AC AAY94454;

XX DT 11-SEP-2000 (first entry)

XX DE Mouse lactadherin protein.

XX KW Human; lactadherin; MGF E8, anti-tumour; immune response;
 XX exosome; dendritic cell.

XX OS Mus sp.

XX FH Key Location/Qualifiers

FT Peptide 1 22

FT Protein /label= Secretion_signal

FT Binding-site 23..426

FT /label= Lactadherin

FT Binding-site 87..89

FT /label= Integrin_binding_site

FT Misc-difference 93..111

FT /note= "encoded by GT"

XX EP1004664-A1

XX PD 31-MAY-2000.

XX PF 24-NOV-1998; 98EP-0492925.

XX PP 24-NOV-1998; 98EP-0492925.

XX PS (INRM) INSEPM INST NAT SANTE & RECH MEDICALE

XX PA (CURI-) INST CURIE.

XX WPI: 2000-352597/31.

XX N-PSDB; AAA27141.

XX Chimeric isolated (human) lactadherin polypeptide that functions as an
 XX adaptor of cross-priming to eliminate pathogenic antigens, e.g. in
 XX tumors -

XX Dislosure, Page 12; 20pp, English.

XX Lactadherin protein was found in exosomes produced by dendritic cells.
 CC The protein is involved in the phagocytosis of particulate antigens by
 CC dendritic cells. Exosomes produced by dendritic cells exposed to
 CC tumour antigens induce potent immune responses. Lactadherin or variants
 CC of it may be used in the mediation of an immune response. Variants of
 CC lactadherin may be used for inhibition and/or stimulation of the
 CC cross priming of antigens and stimulation of the phagocytosis of
 CC antigens by dendritic cells. Compositions derived from lactadherin can

CC also be used to monitor an immune response, more specifically a CTL
 CC (cytotoxic T lymphocyte) response and also to produce CTLs specific
 CC for a selected antigen. The present sequence is the mouse
 CC lactadherin protein.

XX Sequence 426 AA;

Query Match 84.1%; Score 37; DB 21; Length 426;

Best Local Similarity 87.5%; Pred. No. 1.2e+02;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLIAALCG 8

Db 5 RLIAALCG 12

RESULT 5

AAP83149

ID AAP83149 standard; protein; 345 AA.

XX AC AAP83149;

XX DT 27-NOV-1990 (first entry)

XX DE Probe F10 encoded protein of MHC class I of chicken.

XX KW chicken erythrocyte; histocompatibility complex class I antigen;
 XX restriction fragment length polymorphism (RFLP); probe F10.

XX OS Gallus gallus.

XX FH Key Location/Qualifiers

FT Peptide 1..22

FT /label=signal peptide

FT Domain /note= "derived from mature protein"

FT Domain 23..110

FT /label=alpha 1

FT Domain 111..201

FT /label=alpha 2

FT Domain 202..292

FT /label=alpha 3

FT Region 293..334

FT /label=transmembrane

FT Region 335..345

FT /label=cytoplasmic extremity

XX W08809386-A.

XX PD 01-DEC-1988.

XX PF 27-MAY-1988, 88WC FE00270.

XX PP 29-MAY-1987; 97FP-0097577.

XX PA (DVAL/) DE LA VALLEE BS.

XX Aaffray C, Behar G, Billault A, Chausse AM, Guillemot F, Bourlet Y;

XX WPI: 1988-353959/49.

XX N-PSDB; AAN80603.

XX Restriction fragment length polymorphism genotyping test -

XX for chicken, using erythrocyte DNA fragments and hybridisation

XX probe derived from histocompatibility complex antigen.

XX Dislosure, 7 p; French.

XX Probe F10 is used in RFLP analysis of chicken erythrocyte. Isolated
 CC DNA. It corresponds to a MHC class I antigen and was obtained from
 CC the total mRNA of different tissues of chicken.
 CC See also AAN80602 and AAN2129.

XX

SQ Sequence 345 AA;

Query Match 81.8%, Score 36; DB 9; Length 345;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLAALCGA 9
Db 12 llaavcga 19

RESULT 6
AAAY94506
ID AAY94506 standard; Protein; 355 AA.

XX AC AAY94506;

XX DT 20-SEP-2000 (first entry)

XX DE Chicken BFIV21 class I MHC protein.

XX KW Chicken; MHC; major histocompatibility complex; BFIV;

XX KW antisera.

XX OS Gallus gallus

XX FH Key Location/Qualifiers

FT Peptide 1..21 /label= signal_peptide

FT Domain 22..109

FT Domain /label= alpha1_domain

FT Domain 110..200

FT Domain /label= alpha2_domain

FT Domain 201..291

FT Domain /label= alpha3_domain

FT Domain 292..355

FT Modified-site 58 /label= transmembrane_and_cytoplasmic

FT Modified-site 106 /label= N-glycosylation_site

FT Modified-site /label= N-glycosylation_site

XX PN US6075125-A

XX PR 13-JUN-2000

XX PD 09-JUL-1997; 97US-0840719

XX PA 10-JUL-1996; 96US-0021685.

XX PP (USDA) US SEC OF AGRIC.

XX PI Hunt HD, Bacon LD, Fulton JE;

XX PR WPI; 2000-411285/35.

XX DR N-PSDB; AAA48669.

XX PT Producing antisera specific to major histocompatibility complex (MHC)

XX PT proteins in chickens involves administering transfected cells

XX PT expressing heterologous chicken MHC class I protein capable of

XX PT eliciting immune response

XX PS Disclosure; Fig 4, 40pp; English

XX CC The chicken Major Histocompatibility Complex (MHC) B-complex is

XX CC comprised of three classes of loci. Class I was mutated by

XX CC site directed mutagenesis. Transfected cells containing the mutant

XX CC sequence may be generated. The heterologous BFIV protein produced by

XX CC these cells may be used as an immunogen to produce chicken MHC class I

XX CC specific antisera. This antisera may then be used to determine the

XX CC B haplotype of any chicken. BFIV specific antisera may be used to

XX CC determine the B haplotype of chickens with reduced cross reaction

CC with class I and class IV MHC proteins. The present sequence is

CC BFIV21 class I MHC protein.

XX SQ Sequence 355 AA;

Query Match 81.8%, Score 36; DB 21; Length 355;

Best Local Similarity 87.5%; Pred. No. 1.5e+02;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLAALCGA 9
Db 12 llaavcga 19

RESULT 7

AAAY94507

ID AAY94507 standard; Protein; 355 AA.

XX AC AAY94507;

XX DT 20-SEP-2000 (first entry)

XX DE Chicken BFIV12 class I MHC protein.

XX KW Chicken; MHC; major histocompatibility complex; BFIV;

XX KW antisera.

XX OS Gallus gallus.

XX FH Key Location/Qualifiers

FT Peptide 1..21 /label= signal_peptide

FT Misc-difference 3 /note= "encoded by TCG"

FT Domain 22..109

FT Modified-site 58 /label= alpha1_domain

FT Modified-site 106 /label= N-glycosylation_site

FT Domain 110..200 /label= N-glycosylation_site

FT Domain /label= alpha2_domain

FT Domain 201..291 /label= alpha3_domain

FT Domain 292..355

FT /label= transmembrane_and_cytoplasmic

XX PN US6075125-A.

XX PR 13-JUN-2000.

XX PD 09-JUL-1997; 97US-0890719.

XX PA 10-JUL-1996; 96US-0021685.

XX PP (USDA) US SEC OF AGRIC.

XX PI Hunt HD, Bacon LD, Fulton JE;

XX PR WPI; 2000-411285/35.

XX DR N-PSDB; AAA48670

XX PT Producing antisera specific to major histocompatibility complex (MHC)

XX PT proteins in chickens involves administering transfected cells

XX PT expressing heterologous chicken MHC class I protein capable of

XX PT eliciting immune response

XX PS Disclosure; Fig 4; 40pp; English.

XX CC The chicken Major Histocompatibility Complex (MHC) B complex is

XX CC comprised of three classes of loci. Class I was mutated by

XX CC site directed mutagenesis. Transfected cells containing the mutant

CC sequence may be generated. The heterologous BIV protein produced by
 CC these cells may be used as an immunogen to produce chicken MHC class I
 CC specific antisera. This antisera may then be used to determine the
 CC B haplotype of any chicken. BIV specific antisera may be used to
 CC determine the haplotype of chickens with reduced cross reaction
 CC with class I and class IV MHC proteins. The present sequence is
 CC BIV12 class I MHC protein.
 XX
 SQ Sequence 355 AA;

Query Match 81.8%, Score 95, IE 21, Length 355;
 Best Local Similarity 87.5%, Prod. No. 1, 6000;
 Matches 7; Conservative 1; Mismatches 6, Indels 0, Gaps 0;

QY 2 LLAALCGA 9
 Db 12 llaavega 19
 |||||

RESULT 8
 ABG15914
 ID ABG15914 standard; Protein: 365 AA.
 XX
 AC ABG15914;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #15905.
 XX
 KW Human, chromosome mapping, gene mapping, gene therapy; forensic,
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 FN W0200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAP-2001; 2001WO-0508631.
 XX
 PR 31-MAP-2000; 2000US-0540217
 PR 23-AUG-2000; 2000US-0540167.
 XX
 PA (HYSE-) HYSEQ INC
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 WP1: 2001-Pat362/72
 DP N-PSDB: AAS80101.
 XX
 PT New isolated polynucleotide and encoded polypeptide may be useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID No 46273; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity.

CC and to produce other types of data and products dependent on RNA and
 CC amino acid sequences. ABG0610-ABG3077 represent novel human
 CC diagnosed amino acid sequences of the invention
 CC Note: the sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pat_sequences.
 XX
 SQ Sequence 365 AA;

Query Match 79.5%, Score 95, DB 22, Length 365;
 Best Local Similarity 100.0%, Prod. No. 2, 2, 2, 2;
 Matches 7; Conservative 0, Mismatches 0, Indels 0, Gaps 0;

QY 2 LLAALCG 8
 Db 85 llaalcg 91
 |||||

RESULT 9
 AAU40306
 ID AAU40306 standard; Protein: 160 AA.
 XX
 AC AAU40306;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Propionibacterium acnes immunogenic protein #1202.
 XX
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; cervicomyelitis;
 KW arthritis; osteophthalmitis; bone joint, central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological, osteopathic; neutrophilic.
 XX
 OS Propionibacterium acnes.
 XX
 FN W0200181581-A2.
 XX
 PD 01-NOV-2001.
 XX
 PF 20-APR-2001; 2001WO-0512865.
 XX
 PR 21-APR-2002; 2000US-1990477.
 PR 02-JUN-2000; 2000US-208941P.
 PR 07-JUL-2000; 2000US-216747P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Skocky YAW, Peising DH, Mitcham CL, Ward SS, Blattia A;
 PI L'maisonneuve J, Zhang Y, Jan S, Carter D;
 XX
 WP1: 2001-516774/71.
 DP N-PSDB: AAS59511.
 XX
 PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris.
 XX
 PS Example 1; SEQ ID No 1501; 1059pp; English.
 XX
 CC Sequences AAU39105-AAU68017 represent propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated RNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), avelitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence of absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to

CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 160 AA;

Query Match 77.3%, Score 34; DB 22; Length 160;

Best Local Similarity 66.7%, P-val No. 1 5e-02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLAAALCGA 9

II :||||

Db 17 RLAAALCGA 25

RESULT 10

AAU41965

ID AAU41965 standard; Protein; 197 AA.

XX AC AAU41965;

XX AC AAU41965;

XX 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #2861

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181541-A2.

XX 01-NOV-2001.

XX 29-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-166647.

XX 02-JUN-2000; 2000US-208841P.

XX 07-JUL-2000; 2000US-216747P

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Perrina BL, Mirchand IL, Wang SS, Rhatia A;

PI L'Haissonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

XX N-PSDB; AAS59515.

XX Propionibacterium acnes polypeptides and nucleic acids useful for

PT vaccinating against and diagnosing infections, especially useful for

PT treating acne vulgaris.

XX Example 1, SEQ ID No. 130, 131-134, English

XX Sequences AAU3-4195 AAU6-8017 represent propionibacterium acnes immunogenic

CC polypeptides. The proteins and their associated DNA sequences are used in

CC the treatment, prevention and diagnosis of medical conditions caused by

CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,

CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis

CC P. acnes is also involved in infections of bone, joints and the central

CC nervous system, however it is particularly involved in the inflammatory

CC lesions associated with acne vulgaris. A method for detecting the

CC presence or absence of P. acnes in a patient comprises contacting a

CC sample with a binding agent that binds to the proteins of the invention

CC and determining the amount of bound protein in the sample. The

CC polypeptides may be used as antigens in the production of antibodies

CC specific for P. acnes proteins. These antibodies can be used to

CC downregulate expression and activity of P. acnes polypeptides and

CC therefore treat P. acnes infections. The antibodies may also be used as

CC diagnostic agents for determining P. acnes presence, for example, by

CC enzyme linked immunosorbent assay (ELISA).

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 197 AA;

RESULT 11

ABB70578

ID ABB70578 standard; Protein; 466 AA.

XX AC ABB70578;

XX AC ABB70578;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID No 38526.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614159.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers BW;

XX WPI; 2001-616774/75.

XX N-PSDB; ABL14681.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions.

XX Disclosure; SEQ ID No 38526; 21pp. Sequence listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (AB16176, AB16177), expressed DNA

CC sequences (AB16178-AB16179) and the encoded proteins

CC (AHR57/4)-AHR57/2072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 466 AA;

Query Match 77.3%; Score 34; DB 22; Length 466;

Best Local Similarity 75.0%; Prod No. 4 2e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY 1 RLAAACG 8
|||||
DB 122 rllaagc 129

RESULT 12

ABG18818
ID ABG18818 standard; Protein: 1754 AA.

XX AC ABG18818;

XX DI 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #18809.

XX HW Human, chromosome mapping, gene mapping, gene therapy, forensic,
XX KW food supplement, medical imaging, diagnostic, genetic disorder

XX GS Homo sapiens.

XX PN W0200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PP 31-MAR-2001; 2001US-0543217.

XX PR 23-AUG-2001; 2001US-0543217.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-039362/73.

XX DR N-PSDB: AAS83005.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -

XX PS Claim 20, SEQ ID NO 49177, 139pp, English.

XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromo-
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC quantitating a polypeptide in tissue as molecular weight markers and as
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of 3D and 2D arrays dependent on DNA and
XX CC amino acid sequences. AAG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.

XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at http://wipo.int/pat/published_pat_sequences

XX SQ Sequence 1754 AA;

Query Match

77.3%; Score 34; DB 22; Length 1754;

Best Local Similarity 77.8%; Prod No. 1 5e+04;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Caps 0;

QY 1 RLAAACG 9
|||||
DB 1587 rllaalesa 1595

RESULT 13

AA009273
ID AA009273 standard; Protein: 1759 AA.

XX AC AA009273;

XX DI 06-NOV-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 23165.

XX KW Human, cytokine, cell proliferation, cell differentiation, gene therapy,
XX KW various, legume therapy, stem cell growth factor, hematopoiesis,
XX KW tissue, tissue factor, immunomodulatory factor, leukaemia,
XX KW nervous system disorders, arthritis, inflammation.

XX GS Homo sapiens.

XX PN W0200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PP 26 FEB 2001; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX WPI: 2001 514838/56.

XX DR N-PSDB: AA189204.

XX PT Isolated nucleic acids and polypeptides, useful for preventing
XX PT diagnosing and treating e.g. leukaemia, inflammation and immune
XX PT disorders -

XX PS Claim 26; SEQ ID NO 23165; 1399pp - Sequence Listing; English.

XX CC The invention relates to human polynucleotides (AA179041 AA179041) and
XX CC the encoded proteins (AA009273-AA009273) that exhibit activity relating to
XX CC cytokine, cell proliferation or cell differentiation of which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, hematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activity/tissue activity and may be useful in the treatment and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation.

XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at http://wipo.int/pat/published_pat_sequences.

XX SQ Sequence 1759 AA;

Query Match

77.8%; Score 34; DB 22; Length 1759;

Best Local Similarity 77.8%; Prod No. 1 5e+04;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Caps 0;

QY 1 RLAAACG 9
|||||
DB 1590 rllaalesa 1598

RESULT 14
AA885537
ID AA885537 standard; protein: 80 AA.
XX
AC
AA885537;
XX
XX
25-SEP-2001 (first entry)
XX
DE Human secreted protein (clone 13 HEP32).
XX
KW secreted protein; immunosuppressive; antiarthritis; antirheumatic;
KW antiproliferative; "histamine", cardiac; vasodilator; cerebroprotective;
KW neurologic; neuroprotective; antibacterial; vitacide; fungicide; human;
KW ophthalmological; gene therapy.
XX
OS Homo sapiens.
XX
XX
FH Key location/qualifiers
FT Misc-difference 11 /note= "Xaa can be any amino acid"
FT
PN W0200155430-A1.
XX
XX 02-AUG-2001.
XX
DE 17-JAN-2001; 2001W0-US01441.
XX
PR 31-JAN-2000; 2000US-0170065.
PR 04-FEB-2000; 2000US-0170065.
PR 12-SEP-2000; 2000US-0231968.
XX
PA (HUMAN) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Komatsuouis GA, Baker KP, Birse CE, Soppet DR, Gissen HS,
PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Piscella M;
PI Ni J, Rubin SM, Parash SC;
XX
DR WPI: 2691-1762-51.
DR N-PSDB: AAH46947.
XX
XX 17 isolated nucleic acid molecules encoding human secreted proteins,
PI used to preventing, treating or ameliorating a medical condition -
XX
PS Claim 11; Page 454; 482pp; English.
XX
XX The invention provides novel human secreted proteins and polynucleotides
XX encoding them. The secreted proteins can be expressed by standard
XX recombinant methodology. The secreted proteins and polynucleotides are
XX used to prevent, treat or ameliorate a medical condition in e.g. humans,
XX mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They can
XX also be used in diagnosing a pathological condition. The antibodies to
XX the proteins can also be used in alleviating symptoms associated with the
XX disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme
XX linked immunosorbent assays (ELISA). Disorders which are diagnosed or
XX treated include autoimmune diseases e.g. rheumatoid arthritis,
XX hyperproliferative disorders e.g. neoplasms of the breast or liver,
XX cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
XX e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi and
XX ocular disorders e.g. corneal infection. The polypeptides can also be
XX used to aid wound healing and epithelial cell proliferation, to prevent
XX skin aging due to sunburn, to maintain organs before transplantation, for
XX supporting cell culture of primary tissues, to regenerate tissues and in
XX chemotaxis. The polypeptides can also be used as a food additive or
XX preservative to increase or decrease storage capabilities. The present
XX sequence represents a human secreted protein.
XX
SQ Sequence 80 AA;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 3 LAALCGA 9
Db 43 maalca 49
RESULT 15
ABG20209
ID ABG20209 standard; Protein: 106 AA.
XX
AC ABG20209;
XX
XX 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #20200.
XX
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
XX W0200175967-A2.
XX
PD 11-OCT-2001.
XX
DE 40-MAR-2001; 2001W0-US00631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 24-APR-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-630362/73
DR N-PSDB: AAS84396
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PI diagnostics, forensics, gene mapping, identification of mutations
PI responsible for genetic disorders or other traits and to assess
PI biodiversity -
XX
PS Claim 20. SEQ ID NO 50568, 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantifying a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00018:ABG0377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_prt_sequences.
XX
SQ Sequence 106 AA;

Query Match 75.0%; Score 33; DB 22; Length 106;
Best Local Similarity 85.7%; Prod. No. 150402;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 LAALCGA 9
:|||||
Db 82 maalqga 88

Search completed: September 5, 2002, 15:26:33
Job time: 373 sec



GenCore version 4.5
Copyright (c) 1993 2000 Genengen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:28:54 ; Search time 69.76 Seconds
(without alignments)
12.393 Million cell updates/sec

Title: US-09-744-804 41
Perfect score: 44
Sequence: 1 RLAAALCA 9

Scoring table: NCOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 64
Maximum Match 100

Listing first 45 summaries

Database : PIR_71: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	84.1	427	2 J04915	gas protein precursor
2	37	84.1	463	1 A36479	milk fat globule membrane protein - mouse
3	36	81.8	143	2 E21142	hypothetical protein
4	36	81.8	287	2 E83399	hypothetical protein
5	36	81.8	345	1 HLC084	MHC class I histoc
6	36	81.8	355	2 I51306	major histocompati
7	36	81.8	355	2 T28149	MHC class I histoc
8	36	81.8	355	2 T28152	MHC class I histoc
9	35	79.5	230	2 E87446	hypothetical prote
10	34	77.3	171	2 E87418	hypothetical prote
11	34	77.3	376	2 F75503	conserved hypothet
12	34	77.3	457	2 A10955	probable PTS syste
13	34	77.3	1744	1 C4H0	complement C4A pre
14	33	75.0	196	2 A00385	4-methyl-5(8-hydro
15	33	75.0	218	2 F76524	hypothetical prote
16	33	75.0	235	2 T04860	hypothetical prote
17	33	75.0	307	2 T19592	hypothetical prote
18	33	75.0	334	2 T04520	probable nicotinac
19	33	75.0	462	1 S30104	indoleacetamide hy
20	33	75.0	466	2 T39787	probable indoleace
21	33	75.0	467	1 HA33	indoleacetamide hy
22	33	75.0	467	1 QACAG	indoleacetamide hy
23	33	75.0	467	2 A13228	indole acetamide E
24	33	75.0	469	2 T25536	probable integral
25	33	75.0	644	2 A36325	epidermal growth f
26	33	75.0	2161	2 T30342	protein DPM1 Yc
27	33	75.0	4717	2 T11581	hypothetical prote
28	32	72.7	102	2 T28283	hypothetical prote
29	32	72.7	164	1 E10266	sigma E factor reg

30 32 72.7 159 2 D91058
31 32 72.7 159 2 D85903
32 32 72.7 205 2 E25331
33 32 72.7 244 2 E97532
34 32 72.7 244 2 A02751
35 32 72.7 274 2 E21426
36 32 72.7 305 2 T04558
37 32 72.7 313 2 H73777
38 32 72.7 325 2 A01047
39 32 72.7 335 2 E97207
40 32 72.7 335 2 E94224
41 32 72.7 344 2 E84502
42 32 72.7 344 2 E72119
43 32 72.7 355 2 E51152
44 32 72.7 356 2 E75266
45 32 72.7 357 2 E35291

ALIGNMENTS

RESULT 1
JC4915
N: Alternate names: O-acetyl-Gd3 ganglioside
C: Species: Rattus norvegicus (Norway rat)
C: Date: 26 Sep-1996 #sequence_revision: 01-Nov-1996 #text_change: 20-Jun-2000
C: Accession: JC4915
Biochem. Biophys. Res. Commun. 225, 932-938, 1996
A: Title: Cloning and expression of cDNA for O-acetylation of Gd3 ganglioside.
A: Reference number: JC4915, MIM:3637422
A: Accession: JC4915
A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 1-427 <550>
A: Cross references: DDBJ:U04068, EMBL:U04068, F019:U04068, F019:U04068
A: Experimental source: GST cell
A: Comment: This protein is required for the O acetylation of disialoganglioside sial
C: Geneties:
A: Genes: ags
C: Superfamily: milk fat globule protein, discoidin 1 amino terminal homology: EGF hom
F: 1-237 domain, 3-sect sequence #status predicted: EIC
F: 28-60 domain, EGF homology <EGF>
F: 68-197 domain, EGF homology <EGF>
F: 193-367 domain, discoidin 1 amino terminal homology <DM1>
F: 427 domain, discoidin 1 amino terminal homology <DM1>

Query Match 44 191 Score 47 DB 2 Length 427
Best Local Similarity 97.6% Pred. No. 29
Matches 7, Conservative 1, Mismatches 0, Indels 0, Gaps 0

QY 1 RLAAALCA 8
DL 5 RLAAALCA 12

RESULT 2
A36479
milk fat globule membrane protein - mouse
C: Species: Mus musculus (house mouse)
C: Date: 10-Sep-1999 #sequence_revision: 10 Sep 1999 #text_change: 10-Sep-1999
C: Accession: A36479
Biochem. Biophys. Res. Commun. 255, 932-938, 1999
A: Title: cDNA cloning of a mouse mammary epithelial cell surface protein reveals the
A: Reference number: A36479, MIM:3646008
A: Accession: A36479
A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 1-463 <570>
A: Cross references: GB:M38237, NID:q199142, PID:AAA36534.1, PID:q199143

C:Suprafamily: milk fat globule protein; discoidin I amino-terminal homology: PGP homolog
C:Keywords: membrane protein
F:29-69/Domain: EGF homology <E31>
F:68-107/Domain: EGF homology <E32>
F:147-204/Domain: discoidin I amino-terminal homology <E1>
F:167-414/Domain: discoidin I amino-terminal homology <E2>

Query Match 84.1% Score 37; DB 1; Length 463;
Best Local Similarity 87.5%; Pred. No. 31;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0.

QY 1 LLAALGCA 8
|||||
DB 5 LLAALGCA 12

RESULT 3
DB3148
Hypothetical protein PA3986 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: DB3148
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Polger, K.P.; Kas, A.; Lartig, K.; Lim, C.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: AB2450; MUID:20437337
A:Accession: DB3148
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-143 <STG>
A:Cross-references: GB:AE064654; JF:AE064654; FIDN:AA305365; FIDN:AA305365
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3986

Query Match 81.8% Score 36; DB 2; Length 143;
Best Local Similarity 77.8%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLAALGCA 9
|||||
DB 108 LLAALGCA 116

RESULT 4
DB3398
Hypothetical protein PA1977 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: DB3398
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Polger, K.P.; Kas, A.; Lartig, K.; Lim, C.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: AB2950; MUID:20437337
A:Accession: DB3398
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-287 <STG>
A:Cross-references: GB:AE064654; GB:AE064654; FIDN:AA305365; FIDN:AA305365
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1977

Query Match 81.8% Score 36; DB 2; Length 287;
Best Local Similarity 87.5%; Pred. No. 32;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLAALGCA 9
|||||
DB 242 LLAALGCA 249

RESULT 5
DB3148
MHC class I histocompatibility antigen B-F IV alpha chain precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 22-Jun-1999
C:Accession: A45846; S01172
R:Kroemer, G.; Zoorob, R.; Aufray, C.
Immunogenetics 31, 405-409, 1990
A:Title: Structure and expression of a chicken MHC class I gene.
A:Reference number: A45846; MUID:90316612
A:Accession: A45846
A:Molecule type: DNA
A:Residues: 1-345 <KRO>
A:Cross-references: GB:M31012
R:Guillemot, F.; Billault, A.; Pourquie, O.; Behar, G.; Chausse, A.M.; Zoorob, R.; Kr
EMBO J. 7, 2775-2785, 1988
A:Title: A molecular map of the chicken major histocompatibility complex: the class I
A:Reference number: S01172; MUID:89030642
A:Accession: S01172
A:Molecule type: mRNA
A:Residues: 1-345 <GUI>
A:Cross-references: EMBL:X12780; NID:q63089; FIDN:CAA31272 1; PID:q63090
C:Genetics:
A:Map position: 16
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology
C:Keywords: glycoprotein; heterodimer; transmembrane protein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-301/Domain: extracellular #status predicted <EXT>
F:302-324/Domain: transmembrane #status predicted <TM>
F:325-345/Domain: intracellular #status predicted <INT>
F:59,107/Binding site: carbohydrate (ASN) (covalent) #status predicted
F:121-183,221 277/Disulfide bonds: #status predicted

Query Match 81.8% Score 36; DB 1; Length 345;
Best Local Similarity 87.5%; Pred. No. 37;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLAALGCA 9
|||||
DB 12 LLAALGCA 19

RESULT 6
DB3148
Major histocompatibility complex class I glycoprotein haplotype B*01 - chicken
C:Species: Gallus gallus (chicken)
C:Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 21-Jan-2000
C:Accession: I51309
R:Fulton, J.E.; Thacker, E.L.; Bacon, L.D.; Hunt, H.D.
Eur. J. Immunol. 25, 2069-2076, 1995
A:Title: Functional analysis of avian class I (B*01) glycoproteins by epitope mapping
A:Reference number: I51309; MUID:95347411
A:Accession: I51309
A:Status: preliminary; translated from GB:EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-355 <FUI>
A:Cross-references: GB:S78662; NID:q1942200; FIDN:AA04445 1; PID:q1942201
C:Genetics:
A:Gene: B*01
C:Superfamily: class I histocompatibility antigen; immunoglobulin homology
F:213-278/Domain: immunoglobulin homology <IMM>

A:Accession: F75504
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-376 <WHI>
 A:Cross-references: GB:AF001914, GB:AE030513, NID:3458253, PIRN:AAF16140 1, PIR:345829
 C:Genetics:
 A:Gene: DR0560
 A:Map position: 1

Query Match 77.3% Score 34; DB 2; Length 376;
 Best Local Similarity 88.9% Pred. No. 92;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Caps 0

QY 1 RLAAALCGA 9
 |||||
 DB 48 RLAAALCGA 46

RESULT 12
 A10955
 probable PTS system precursor (imported) - Salmonella enterica subsp. enterica serovar Typhimurium
 C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: this species has also been called Salmonella typhi
 C:Date: 04-Nov-2001 *sequence_revision: 09-Nov-2001 *text_change: 09-Nov-2001
 C:Accession: A10955
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, R.R.; Pickard, D.; Wain, J.; Churcher, C.; White, N.; Davies, P.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, J.; Mouton, S.; O'Garra, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Ouali, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
 A:Reference number: AR0502; PMID:11277526
 A:Accession: A10955
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-457 <PAR>
 A:Cross-references: GB:A151382, FIDN:CAJ03141 1, PIR:34584776, GSPDB:GN00176
 C:Genetics:
 A:Gene: STY4925

Query Match 77.3% Score 34; DB 2; Length 457;
 Best Local Similarity 75.0% Pred. No. 116;02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Caps 0

QY 2 LLAAALCGA 9
 |||||
 DB 204 LLAAALCGA 211

RESULT 13
 C4HD
 complement C4A precursor [validated] - human
 N:Contains: classical complement pathway C3/C5 convertase (P1 3 4 21 43) C4a subunit, C4b
 C:Species: Homo sapiens (man)
 C:Date: 25 Feb 1995 *sequence_revision: 03-Aug-1996 *text_change: 28 Dec 2000
 C:Accession: 156095; A29177; A90845; A19311, A23337, S22806, A7205, A23337
 R:Yu, C.Y.
 J. Immunol. 145, 1057-1066, 1991
 A:Title: The complete exon-intron structure of a human complement component C4A gene, EN
 A:Reference number: 156095; PMID:91108036
 A:Accession: 156095
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1744 <PFS>
 A:Cross-references: GB:M56405; NID:q17672; PIRN:AAAC1855.1, PIR:q17674
 R:Belt, K.T.; Yu, C.Y.; Carroll, M.C.; Porter, R.R.
 Immunogenetics 21, 173-180, 1985
 A:Title: Polymorphism of human complement component C4
 A:Reference number: A29177; PMID:8515629
 A:Accession: A29177

A:Molecule type: DNA
 A:Residues: 1-22 <BEL>
 A:Cross-references: GB:M14823
 A:Accession: B29177
 A:Molecule type: DNA
 A:Residues: 1056-1225 <BE3>
 A:Cross-references: GB:M14824; NID:q179675; PIRN:AAAS2292.1; PIR:q553210
 R:Belt, K.T.; Carroll, M.C.; Porter, R.R.
 Cell 36, 907-914, 1984
 A:Title: The structural basis of the multiple forms of human complement component C4.
 A:Reference number: A90845; PMID:84156544
 A:Accession: A90845
 A:Molecule type: mRNA
 A:Residues: 29-345; S' 349-417; A' 419-725; P' 727-1290; S' 1292-1295; S' 1297-1418; 1
 R:Carroll, M.C.; Porter, R.R.
 Proc. Natl. Acad. Sci. U.S.A. 80, 264-267, 1983
 A:Title: Cloning of a human complement component C4 gene.
 A:Reference number: A19311; PMID:83117835
 A:Accession: A19311
 A:Molecule type: mRNA
 A:Residues: 1195-1285; S' 1287-1294 <CAK>
 A:Cross-references: GB:V05032; GB:205083; NID:q30010; PIRN:CAA3760.1; PIR:q1345040
 J:Moore, K.E.; Gorski, J.P.; Huq, T.E.
 J. Biol. Chem. 256, 8685-8692, 1981
 A:Title: Complete primary structure of human C4a anaphylatoxin.
 A:Reference number: A92337; PMID:81264286
 A:Accession: A92337
 A:Molecule type: protein
 A:Residues: 680-725; P' 728-756 <WVS>
 R:Hessing, M.; Van't Veer, C.; Hacking, T.M.; Bouma, B.N.; Iwanaga, S.
 FEBS Lett. 271, 131-136, 1990
 A:Title: Importance of the alpha(3)-fragment of complement C4 for the binding with C4
 A:Reference number: S12866; PMID:91032049
 A:Accession: S12866
 A:Molecule type: protein
 A:Residues: 757; X' 759-771; 980-990 <HES>
 R:Campbell, R.D.; Gagnon, J.; Porter, R.R.
 Biochem. J. 194, 559-570, 1981
 A:Title: Amino acid sequence around the thiol and reactive acyl groups of human comp
 A:Reference number: A17255; PMID:62192029
 A:Accession: A17255
 A:Molecule type: protein
 A:Residues: 957-1012; E' 1014-1044 <CAM>
 R:Chakravarti, D.N.; Campbell, R.D.; Porter, R.R.
 Mol. Immunol. 24, 1187-1197, 1987
 A:Title: The chemical structure of the C4d fragment of the human complement component
 A:Reference number: A32335; PMID:88094444
 A:Accession: A32335
 A:Molecule type: protein
 A:Residues: 957-1012; E' 1014-1044 <CAM>
 R:Chakravarti, D.N.; Campbell, R.D.; Porter, R.R.
 Mol. Immunol. 24, 1187-1197, 1987
 A:Title: The chemical structure of the C4d fragment of the human complement component
 A:Reference number: A32335; PMID:88094444
 A:Accession: A32335
 A:Molecule type: protein
 A:Residues: 957-1012; E' 1014-1044 <CAM>
 R:Chakravarti, D.N.; Campbell, R.D.; Porter, R.R.
 Mol. Immunol. 24, 1187-1197, 1987
 A:Title: Amino acid sequence of a polymorphic segment from fragment C4d of human comp
 A:Reference number: A27600; PMID:83158189
 A:Accession: A27600
 A:Molecule type: protein
 A:Residues: 1199-1270; V' 1272-1290; V' 1301-1304 <BE2>
 R:Whitehead, A.S.; Goldberger, G.; Woods, D.F.; Markham, A.F.; Colten, B.R.
 Proc. Natl. Acad. Sci. U.S.A. 80, 5487-5491, 1983
 A:Title: Use of a cDNA clone for the fourth component of human complement (C4) for an
 A:Reference number: 158991; PMID:8329999
 A:Accession: 158991
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1448-1474 <PE2>
 A:Cross-references: GB:K0833; NID:q18772; PIRN:AAAG6229.1; PIR:q18773
 R:Sargent, C.A.; Anderson, M.J.; Hsieh, S.L.; Kendall, P.; Gomez-Fescobar, N.; Campbell
 Hum. Mol. Genet. 3, 481-488, 1994
 A:Title: Characterisation of the novel gene G11 lying adjacent to the complement C4A
 A:Reference number: 137396; PMID:94282044
 A:Accession: 137396

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:31.43, Search time 33.99 seconds
(without alignments)
10.252 Million cell updates/sec

Title: US-09-744-804-41
Perfect score: 44
Sequence: 1 KLLAALGCA 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	44	100.0	387	1 MFGM_HUMAN	Q08431 homo sapien
2	37	84.1	427	1 MFGM_FAT	P20490 rattus norv
3	37	84.1	463	1 MFGM_MOUSE	P21956 mus musculus
4	36	81.8	287	1 GLPS_PSPAP	P29370 pseudomonas
5	36	81.8	345	1 HALF_CHICK	P15979 gallus gall
6	34	77.4	879	1 MANB_HUMAN	Q00462 homo sapien
7	34	77.3	1744	1 COA_HUMAN	P01028 homo sapien
8	33	75.0	245	1 CBIL_SALTY	Q05594 salmonella
9	33	75.0	267	1 THIM_PASMU	P57931 pasteurella
10	33	75.0	334	1 COB1_PHOCA	Q52679 rhodobacter
11	33	75.0	462	1 HVT1_AGRVI	Q04557 agrobacteri
12	33	75.0	466	1 HYIN_AGRPH	Q09102 agrobacteri
13	33	75.0	467	1 HVT2_AGRVI	P25016 agrobacteri
14	33	72.7	159	1 RSEC_ECOLI	P03868 agrobacteri
15	32	72.7	313	1 CORO_MYCTO	P46187 escherichia
16	32	72.7	325	1 SVK3_SALTY	Q10518 mycobacteri
17	32	72.7	357	1 C3BT_STRCP	Q92312 salmonella
18	32	72.7	357	1 C3BT_STRCP	Q92312 streptomyce
19	32	72.7	363	1 MBS1_SACDO	P41905 saccharomyc
20	32	72.7	384	1 THIL_MYCTE	Q56225 mycobacteri
21	32	72.7	393	1 THIL_MYCLE	P46707 mycobacteri
22	32	72.7	422	1 U193_HUMAN	Q08431 homo sapien
23	32	72.7	422	1 U193_HUMAN	Q08431 mus musculus
24	32	72.7	422	1 U193_HUMAN	Q08431 rattus norv
25	32	72.7	444	1 TRFD_ARATH	Q02166 arabidopsis
26	32	72.7	455	1 SYN_MYCTH	P75521 mycoplasma
27	32	72.7	456	1 SYN_MYCCE	P47359 mycoplasma
28	32	72.7	531	1 LAT2_MOUSE	Q94629 mus musculus
29	32	72.7	533	1 LAT2_FAT	Q94629 rattus norv
30	32	72.7	535	1 LAT2_HUMAN	Q94629 homo sapien
31	32	72.7	1876	1 PSRA_MYCTO	Q10477 mycobacteri
32	32	72.7	2280	1 GCAC_SCHPO	P28826 schistosom
33	31	70.5	196	1 THIL_ECOLI	Q46948 escherichia

RESULT 1

ID	MFGM_HUMAN	STANDARD	PRT	387 AA.
AC	Q08431			
DT	01-OCT-1996	(Ref. 34, Created)		
DT	01-NOV-1997	(Ref. 35, Last sequence update)		
DT	16-DEC-2001	(Ref. 40, Last annotation update)		
DE	Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-EG8) (HMF3)			
DE	(arrest epithelial antigen BA46) (MFGM) [Contains: Modin].			
GN	MFG8.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
ON	(1)			
RP	SEQUENCE FROM N.A.			
EC	TISSUE Breast, and breast carcinoma;			
EX	MEDLINE:96213908; PubMed:863264;			
EA	Couto J.P., Taylor M.R., Godwin S.G., Ceriani R.L., Peterson J.A.;			
EA	"Cloning and sequence analysis of human breast epithelial antigen			
EA	BA46 reveals an epithelial cell adhesion sequence presented on an epidermal			
EA	growth factor-like domain.";			
EL	DNA Cell Biol. 15:281-286(1996)			
RN	[2]			
RN	SEQUENCE OF 170-387 FROM N.A.			
RC	TISSUE Mammary gland;			
EX	MEDLINE 91371361; PubMed 1909932;			
EA	Faracca D., Peterson J.A., Urrea P., Kunyoshi J., Bistrain A.M.,			
EA	Ceriani R.L.;			
EA	"A 16,000 human milk fat globule protein that is highly expressed			
EA	in human breast tumors contains factor VIII-like domains.";			
EA	Cancer Res. 51:4994-4998(1991).			
RN	[3]			
RC	PARTIAL SEQUENCE, AND CHARACTERIZATION.			
RC	TISSUE Milk;			
EX	MEDLINE 98191921; PubMed 9535276;			
EA	Glutirida M.G., Cavalletto M., Giunta C., Conti A.,			
EA	Gebreyes Zimmehman J.,			
EA	"Isolation and characterization of full and truncated forms of human			
EA	breast car-leukemia protein BA46 from human milk fat globule membranes.";			
EA	J. Protein Chem. 17:143-148(1998).			
RN	[4]			
RN	SEQUENCE OF 170-387, AND CHARACTERIZATION OF MODIN.			
EX	MEDLINE 99342876; PubMed=10411933;			
EA	Boonprast K., Nussland J., Slattery K., Westmark G.T., Mochizuki G.,			
EA	Tjundt L., Bensted C., Engstrom U., Westmark P.,			
EA	"Medin an integral fragment of acetic smooth muscle cell-produced			
EA	Lactadherin forms the most common human amyloid.";			
EA	Proc. Natl. Acad. Sci. U.S.A. 96:8629-8634(1999).			
RN	[5]			
RC	CHARACTERIZATION.			
EX	MEDLINE 97497285; PubMed 936029;			
EA	Taylor M.R., Couto J.P., Scallan C.D., Ceriani R.L., Peterson J.A.;			
EA	"Lactadherin (formerly BA46), a membrane-associated glycoprotein			
EA	expressed in human milk and breast carcinomas, promotes A14 G4-Asp			
EA	(RGP) dependent cell adhesion.";			

ALIGNMENTS

RL DNA Cell Biol 16:861-869(1997)
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. BINDS
 CC SPECIFICALLY TO PHATININS AND INHIBITS ITS REPLICALLIN.
 CC -1- FUNCTION: MEDIN IS THE MAIN CONSTITUENT OF AORTIC MEDIAL AMYLOID.
 CC -1- SURCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: MAMMARY EPITHELIAL CELL SURFACES AND AORTIC
 CC MEDIA. OVEREXPRESSED IN SEVERAL CAPCINOMAS.
 CC -1- PTM: MEDIN HAS A RAGGED N-TERMINUS WITH MINOR SPECIES STARTING AT
 CC AMINO ACID 264 AND 273.
 CC -1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -----
 CC This SWISS-prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: U58516; AAC50549.1; -
 DR EMBL: S56151; AAB19771.1; -
 DR MIM: 602281; -
 DR InterPro: IPR000561; EGF-like
 DR InterPro: IPR000421; FAS8_C
 DR Pfam: PF00008; EGF_1
 DR Pfam: PF00754; F5_P8_Type_C_2
 DR SMART: SM00181; EGF_1
 DR SMART: SM00231; FAS8C_2
 DR PROSITE: PS00022; EGF_1; 1
 DR PROSITE: PS01186; EGF_2; 1
 DR PROSITE: PS01285; FAS8C_1; 2
 DR PROSITE: PS01284; FAS8C_2; 2
 KW Signal: glycoprotein; Milk; Repeat: EGF-like domain; Amyloid
 FT SIGNAL 1 23
 FT CHAIN 24 387 LACTADHERIN, STAND, EGF FORM
 FT CHAIN 232 387 LACTADHERIN, STAND, EGF FORM
 FT CHAIN 268 317 MEDIN
 FT DOMAIN 24 67 EGF-LIKE
 FT DOMAIN 70 225 F5/8 TYPE C 1
 FT DOMAIN 230 387 F5/8 TYPE C 2
 FT SITE 46 48 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 27 38 BY SIMILARITY.
 FT DISULFID 32 55 BY SIMILARITY.
 FT DISULFID 57 66 BY SIMILARITY.
 FT DISULFID 70 225 BY SIMILARITY.
 FT DISULFID 212 216 BY SIMILARITY.
 FT DISULFID 230 387 BY SIMILARITY.
 FT CARBOHYD 338 348 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 325 325 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 324 324 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 350 350 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 387 AA; 43123 MW; 2EE6571DE083782D CRC64;

Query Match 100.0%; Score 44; DB 1; Length 387;
 Best Local Similarity 100.0%; Pred No 0.57;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PIIAALCG 9
 DB 5 PIIAALCG 13

RESULT 2

MPGM_RAT
 ID MPGM_RAT STANDARD, PRT, 427 AA.
 AC P70490;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Lactadherin precursor (Milk fat globule-EGF factor 8) (MFG-E8) (O-
 DE acetyl GD3 ganglioside synthase) (AGS) (MFGM).

GN MFGF8 OR AGS
 OS Ratius norvegicus (Rat).
 CA Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A
 RC TISSUE=Brain;
 PX MEXTRNE-06274422; PubMed-8780713;
 PA Gupta K., Nara K., Watanabe Y., Kohno K., Tai T., Sanai Y.;
 RT "Cloning and expression of cDNA for G acetylation of GD3
 ganglioside.";
 CC -1- FUNCTION: MAY BE INVOLVED IN PHOSPHOLIPID BINDING. SEEMS TO
 CC PARTICIPATE IN THE G-ACETYLATION OF GD3 GANGLIOSIDE SIALIC ACID.
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: SPLEEN, LUNG, HEART, BRAIN AND MUSCLE.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 F5/8 TYPE C DOMAINS.
 CC -----
 CC This SWISS-prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: D84068; BAB12210.1; -
 DR HSSP: P00740; 11XA.
 DR InterPro: IPR000561; EGF-like
 DR InterPro: IPR001438; EGF-11.
 DR InterPro: IPR000421; FAS8_C
 DR Pfam: PF00008; EGF_2
 DR Pfam: PF00754; F5_P8_Type_C_2
 DR PRINTS: PR00181; EGF_2
 DR SMART: SM00181; EGF_2
 DR SMART: SM00231; FAS8C_2
 DR PROSITE: PS00022; EGF_1; 2
 DR PROSITE: PS01186; EGF_2; 2
 DR PROSITE: PS01285; FAS8C_1; 2
 DR PROSITE: PS01284; FAS8C_2; 2
 DR PROSITE: PS01286; FAS8C_2; 2
 KW Signal: glycoprotein; Repeat: EGF-like domain; Milk
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 427 LACTADHERIN.
 FT DOMAIN 24 61 EGF-LIKE 1.
 FT DOMAIN 64 108 EGF-LIKE 2.
 FT DOMAIN 111 267 F5/8 TYPE C 1.
 FT DOMAIN 273 427 F5/8 TYPE C 2.
 FT DISULFID 28 34 BY SIMILARITY.
 FT DISULFID 33 49 BY SIMILARITY.
 FT DISULFID 51 60 BY SIMILARITY.
 FT DISULFID 58 79 BY SIMILARITY.
 FT DISULFID 73 96 BY SIMILARITY.
 FT DISULFID 98 107 BY SIMILARITY.
 FT DISULFID 111 267 BY SIMILARITY.
 FT DISULFID 254 258 BY SIMILARITY.
 FT DISULFID 272 427 BY SIMILARITY.
 FT SITE 87 89 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 61 61 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 230 230 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 280 280 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 390 390 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 427 AA; 47413 MW; EA8C631F3EE6047 CRC64;

Query Match 84.1%; Score 27; DB 1; Length 427;

Best Local Similarity 87.5%; Pred. No. 11;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PIIAALCG 8

DB 5 RVLAALCG 12

KINASES

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcements/> or send an email to license@isb-sib.ch).

EMBL: AE004624; AAC05465 1; -

EMBL: M60805; AAA25828.1; -

DR InterPro: IPR000620; I0F6.

DR Pfam: PF00892; DPF6; 2.

KW Sensory transduction; transferase; kinase; phosphorylation;

CS Complete proteome.

SEQUENCE 287 AA; 25653 MW; 134771; 2.466 cAa;

Query Match 81.8%; Score 36; DB 1; Length 287;

Best Local Similarity 87.5%; Pred. No. 12;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0.

QY 2 LLAALCGA 9

DB 242 LLAALCGA 249

RESULT 5

HAIF_CHICK
 ID HAIF_CHICK STANDARD; PRT; 345 AA.
 AC P15979;
 DT 01-APR-1990 (Rel. 14; Last sequence update)
 DT 15-JUL-1999 (Rel. 38; Last annotation update)
 DE Class I histocompatibility antigen, F10 alpha chain precursor (B F
 DE histocompatibility F10 antigen) (B-F-beta-IV) (B12).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 OC Gallus.
 OX NCBI_TaxID:9031;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN-R12; TISSUE Liver;
 RX MEDLINE:89030642; PubMed:3141149;
 RA Guillemot F., Rillaud A., Bourguie O., Behar G., Chausse A.M.,
 RA Zorob R., Kreibich G., Auffray C.,
 RT "A molecular map of the chicken major histocompatibility complex: the
 RT class II beta genes are closely linked to the class I genes and the
 RT nucleolar organizer.";
 RL EMBL J. 7:2775-2785(1988).
 RN 121
 RP SEQUENCE FROM N.A.
 RX MEDLINE:90316612; PubMed:2470087.
 RA Kroemer G., Zorob R., Auffray C.,
 RT "Structure and expression of a chicken MHC class I gene.";
 RL Immunogenetics 31:405-409(1990)
 CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO
 CC THE IMMUNE SYSTEM.
 CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
 CC MICROGLOBULIN).

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcements/> or send an email to license@isb-sib.ch).

EMBL: X12780; CAA31272 1; -

EMBL: M31012; AAA48947.1; -

DR PIR: S01172; HLCHR4.
 DR PIR: A45846; A45846.
 DR HSSP: P03989; ILSA.
 DR InterPro: IPR003006; Iq_MHC.
 DR InterPro: IPR003597; Iq_c1.
 DR InterPro: IPR001039; MHC_1.
 DR Pfam: PF00047; Iq; 1.
 DR Pfam: PF00124; MHC_1; 1.
 DR ProDom: PD000050; MHC_1; 1.
 DR SMART: SM00407; TGC1; 1.
 DR PROSITE: PS00290; Iq_MHC; 1.
 KW MHC 1; Transmembrane; Glycoprotein; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 345
 FT CLASS I HISTOCOMPATIBILITY ANTIGEN, F10
 FT ALPHA CHAIN.
 FT EXTRACELLULAR ALPHA-1.
 FT EXTRA-CELLULAR ALPHA-2.
 FT EXTRACELLULAR ALPHA-3.
 FT CONNECTING PEPTIDE.
 FT CYTOPLASMIC.
 FT BY SIMILARITY.
 FT N-LINKED (GLCNAC...) (POTENTIAL).
 FT N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 339 339 P > P08603 (IN REF. 2).
 SQ SEQUENCE 345 AA; 38246 MW; 4300CF8091B69A4 CRC64;

Query Match 81.8%; Score 36; DB 1; Length 345;

Best Local Similarity 87.5%; Pred. No. 15;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLAALCGA 9

DB 12 LLAALCGA 19

RESULT 6

MANB_HUMAN
 ID MANB_HUMAN STANDARD; PRT; 879 AA.
 AC 000462; Q9NYX9; Q96BC3;
 DT 15-JUL-1998 (Rel. 36; Created)
 DT 01-MAR-2002 (Rel. 41; Last sequence update)
 DT 01-MAR-2002 (Rel. 41; Last annotation update)
 DE beta-mannosidase precursor (EC 3.2.1.25) (Mannase) (Mannase).
 GN MANBA
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN 111
 RP SEQUENCE FROM N.A.
 RX MEDLINE:98045009; PubMed:9384606;
 RA Aikawa A.H., Kraeger S.A., Leppardt J.E., Marck M., Kleijer W.J.,
 RA Frederici K.H.;
 RT "Human beta mannosidase cDNA characterization and first identification
 RT of a mutation associated with human beta-mannosidosis.";
 RL Hum Mol Genet 7:75-83(1998)
 RN 121
 RP SEQUENCE FROM N.A.
 RA Chang H.-M., Tsai S.-F.;
 RT "Genome sequencing of the chromosome 4q region implicated in human
 RT hepatocellular carcinoma pathogenesis.";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN 131
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RA Strausberg R.;
 RL Submitted (oct-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: EXOGYLOSIDASE THAT CLEAVES THE SINGLE BETA-LINKED
 CC MANNOSE RESIDUE FROM THE NON-REDUCING END OF ALL N-LINKED
 CC GLYCOPROTEIN OLIGOSACCHARIDES.

RP VARIANT C4A6 ALLOTYPE.
 RA MEDLINE:92242905; PubMed:1573268;
 RA Anderson M.J., Milner C.M., Cotton G.H., Campbell R.D.;
 RA The coding sequence of the hemolytically inactive C4A6 allotype of
 RA human complement component C4 reveals that a single arginine to
 RA tryptophan substitution at beta-chain residue 458 is the likely cause
 RA of the defect.*
 RL J. Immunol. 148:3795-3802(1992).
 CC -!- FUNCTION: C4 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
 CC CLASSICAL PATHWAY OF THE COMPLEMENT SYSTEM. IT IS PROCESS-ED BY
 CC ACTIVATED C1 WHICH REMOVS FROM THE ALPHA CHAIN THE C4A
 CC ANAPHYLATOXIN.
 CC -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C4,
 CC C4A ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
 CC INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
 CC PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
 CC BASOPHILIC LEUKOCYTES.
 CC -!- SUBUNIT: THIS PROTEIN IS SYNTHESIZED AS A SINGLE-CHAIN PRECURSOR
 CC AND, PRIOR TO SPLITTING, IS ENZYMATICALLY CLEAVED TO FORM A TRIMER
 CC OF NONIDENTICAL CHAINS (ALPHA, BETA, AND GAMMA).
 CC -!- POLYMORPHISM: HUMAN COMPLEMENT COMPONENT C4 IS POLYMORPHIC WITH AT
 CC LEAST TWO LOCUS C4A AND C4B. 13 ALLELES OF C4A AND 22 ALLELES OF
 CC C4B HAVE BEEN DETECTED. THE ALLELE SHOWN HERE IS C4A4
 CC -!- POLYMORPHISM: THE C4A ALLELES CARRY THE RHOOD GROUP POTPERS WHITE
 CC THE C4B ALLELES CARRY THE RHOOD GROUP CHINDO
 CC -!- DISEASE: THE C4A6 ALLOTYPE IS TOTALLY DEFICIENT IN HEMOLYTIC
 CC ACTIVITY.
 CC -!- MISCELLANEOUS: C4A ALLOTYPES REACT MORE RAPIDLY WITH THE AMINO
 CC GROUP OF PEPTIDE ANTIGENS WHILE C4B ALLOTYPES REACT MORE RAPIDLY
 CC WITH THE HYDROXYL GROUP OF CARBOHYDRATE ANTIGENS.
 CC -!- MISCELLANEOUS: C4 IS A MAJOR HISTOCOMPATIBILITY COMPLEX CLASS-III
 CC PROTEIN.
 CC -!- SIMILARITY: TO C3, C5 AND ALPHA-2-MACROGLOBULIN.
 CC -!- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
 CC
 CC This SWISS PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: K02403; AAH59537.1; -;
 DR EMBL: K02404; AAA59651.1; -;
 DR EMBL: M59815; AAA51855.1; -;
 DR EMBL: M59816; AAA51855.1; -; (M1NFD)
 DR EMBL: U24578; AAA99717.1; -;
 DR EMBL: AF019413; AA867980.1; -;
 DR EMBL: AL049547; CAB89402.1; -;
 DR EMBL: M14823; AAA35617.1; -;
 DR EMBL: M14824; AAA52292.1; -;
 DR EMBL: X77491; CAA54627.1; -;
 DR PIR: A01262; C4HU.
 DR PIR: A17265; A17265.
 DR PIR: A29177; A29177.
 DR PIR: B20807; B20807.
 DR HSSP: P01031; 1KJS.
 DR SWISS-2DPAGE: P01028; HUMAN.
 DR MIM: 120810; -;
 DR MIM: 120820; -;
 DR MIM: 120790; -;
 DR InterPro: IPR002890; AZM_N.
 DR InterPro: IPR001599; Alpha_2_macroglabin.
 DR InterPro: IPR009020; Anaphylatoxin.
 DR InterPro: IPR001840; Anaphylatoxin
 DR InterPro: IPR001134; Nottin_C
 DR Pfam: PF00207; AZM; 1.
 DR Pfam: PF01835; AZM_N; 1.
 DR Pfam: PF01821; ANATO; 1.
 DR Pfam: PF01754; NTP; 1.
 DR PRINTS: PR00004; ANAPHYLATOXN
 DR ProDom: PD00264; Anaphylatoxin 1

DR SMART: SM00104; ANATO; 1.
 DR PROSITE: PS0477; ALPHA_2_MACROGLABULIN; 1.
 DR PROSITE: PS01177; ANAPHYLATOXIN_1; 1.
 DE PROSITE: PS01178; ANAPHYLATOXIN_2; 1.
 KW Complement pathway; Plasma; Glycoprotein; Sulfation; Signal;
 KW Inflammatory response; Polymorphism; Disease mutation;
 KW Blood group antigen.
 FT SIGNAL 1 19
 FT CHAIN 20 675
 FT PROPEP 576 679
 FT CHAIN 680 1446
 FT PROPEP 1447 1453
 FT CHAIN 1454 1744
 FT PEPTIDE 680 756
 FT DOMAIN 702 736
 FT DISULFID 702 728
 FT DISULFID 703 735
 FT DISULFID 716 736
 FT HD10LEST 1010 1013
 FT MOD_RES 1417 1417
 FT MOD_RES 1420 1420
 FT MOD_RES 1422 1422
 FT CARECHIB 326 326
 FT CAPROCHIB 862 862
 FT CAPROCHIB 1329 1328
 FT CARBOHYD 1391 1391
 FT VARIANT 477 477
 FT
 FT VARIANT 725 725
 FT
 FT VARIANT 1073 1073
 FT
 Query Match 77.3%; Score 34; DB 1; Length 1744;
 Best Local Similarity 77.8%; Pred. No. 1.3e+02;
 Matches 7; Conservative 9; Mismatches 2; Labels 0; Gaps 0;
 QY 1 RLAAALAGA 9
 DB 1577 RELATLCSA 1585
 III III I
 DB 1577 RELATLCSA 1585
 RESULT 8
 CBIM_SALTY
 ID CBIM_SALTY STANDARD; PRI: 245 AA.
 AC Q05594;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 31-JUN-1994 (Rel. 23, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE CBIM protein.
 GN CBIM OR STM2023.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria, gamma subdivision, Enterobacteriaceae;
 OC Salmonella.
 OC NCRI_TaxID:602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2;
 RX MEDLINE 93273696; PubMed=8501034;
 RA Roth J.P., Lawrence J.G., Rubenfield M., Kieffer-Bloquins S.,
 RA Church G.M.;
 RA "Characterization of the cobalamin (vitamin B12) biosynthetic genes
 RA of Salmonella typhimurium.*"
 RL J. Bacteriol. 175:3303-3316(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT3 / SCSCL412 / ATCC 700720;
 RX MEDLINE 21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Fulwiler S., Aki J., Baute M., De P., Hon S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Griwal N., Mulvaney E.,
 RA Ryan F., Sun H., Flores L., Miller W., Steneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium

RT LT2.?
 RL Nature 413:852 856(2001).
 CC -1- FUNCTION: NOT KNOWN.
 CC -1- PATHWAY: COBALAMIN BIOSYNTHESIS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ist.sib.ch/unicoln/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: 112066; AAA27364.1; -.
 DR EMBL: AF008789; AAL20927.1; -.
 DR StyGene: SG10045; cbim.
 DR InterPro: IPR002751; Cbim.
 DR Pfam: pf01891; Cbim; 1
 DR ProDom: PD005331; Cbim; 1
 KW Cobalamin biosynthesis. Complete proteome.
 SC SP000044 245 AA; 28245 MW; C710B906B85E971 CRC64;
 Query Match 75.0%; Score 33; DB 1; Length 245;
 Best Local Similarity 75.0%; Pred. No. 38;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 3;
 QY 1 RLAAALCA 9
 Db 72 KVLALCA 80
 RESULT 9
 ID THIM_PASMO STANDARD; PRT; 267 AA.
 AC P57931;
 DT 16-OCT-2001 (Rel. 40, Created)
 DI 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hydroxyethylthiazole kinase (EC 2.7.1.76) (4-methyl-5-keto
 DE hydroxyethylthiazole kinase) (thz kinase) (thz kinase).
 GN THIM OR PM1262.
 OS Pasteurella multocida.
 CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales.
 CC Pasteurella.
 CC NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 FX MPELINE 2145966; PubMed 11248100;
 RA May B.J., Zhang C., Li L., Paustian M., Whittam T.S., Kaper V.,
 PT "Complete genomic sequence of Pasteurella multocida PM70.";
 PL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001)
 CC -1- CATALYTIC ACTIVITY: ADP + 4-methyl-5-(2-hydroxyethyl)-thiazole =
 CC ADP + 4-methyl-5-(2-phosphoethyl)-thiazole.
 CC -1- COFACTOR: MAGNESIUM (BY SIMILARITY).
 CC -1- PATHWAY: THIAMINE BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE THZ KINASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ist.sib.ch/unicoln/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: A006165; AAA0044.1; -.
 DR EMBL: AF560417; HYTHY_Kinase.
 DR Pfam: PF02110; HK; 1.
 DR PRIN2: PF01999; HYTHY_KINASE.
 KW Thiamine biosynthesis. Transferase, Kinase, Air binding, Manganese,
 KW Complete proteome.

FT METAL 91 91 MAGNESIUM (BY SIMILARITY).
 FT METAL 123 123 MAGNESIUM (BY SIMILARITY).
 FT ACT_SITE 194 194 BASE (BY SIMILARITY).
 SQ SEQUENCE 267 AA; 28245 MW; C710B906B85E971 CRC64;
 Query Match 75.0%; Score 33; DB 1; Length 267;
 Best Local Similarity 75.0%; Pred. No. 41;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 2 LLAAGTZA 2
 Db 195 LLSAVCA 202
 RESULT 10
 ID COBT_RHCCA STANDARD; PRT; 334 AA.
 AC Q52679; Q68087;
 DT 16-OCT-2001 (Rel. 40, Created)
 DI 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyl transferase
 DE (EC 2.4.2.21) (NBT-PRT PRT) (N1-alpha-phosphoribosyltransferase).
 GN COBT OR COBT
 OS Rhodobacter capsulatus (Rhodospirillum rubrum).
 CC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
 CC Rhodobacter.
 CC NCBI_TaxID=1061;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SB1003 / SI 10015;
 RX MEDLINE:97404404; PubMed 9256491;
 RA Vreck C., Fuchs V., Maltsev N., Fares J., Haselkorn R., Paustein M.,
 RT "Sequence of a 185-kb segment of the chromosome of Rhodobacter
 RT capsulatus SB1003.";
 PL Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).
 RN [2]
 RP SEQUENCE OF 1-116 FROM N.A.
 RC STRAIN=ATCC 33302 / B10;
 RX WOIHP 6546277; PubMed-7635831;
 RA Pollack M., King G.;
 RT "Identification and sequence analysis of genes involved in late steps
 RT of cobalamin (vitamin B12) synthesis in Rhodobacter capsulatus.";
 PL J. Bacteriol. 177:4481-4487(1995).
 CC -1- FUNCTION: CATALYZES THE SYNTHESIS OF ALPHA-RIBAZOLE-5'-PHOSPHATE
 CC FROM NICOTINATE MONONUCLEOTIDE (NAMN) AND 5,6-
 CC DIMETHYLBENZIMIDAZOLE (DMB).
 CC -1- CATALYTIC ACTIVITY: Beta-nicotinate D-ribonucleotide +
 CC 5-methyl-5,6-dimethylbenzimidazole diphosphate + N1-(5-phospho-alpha-D-
 CC ribosyl)-5,6-dimethylbenzimidazole.
 CC -1- PATHWAY: COBALAMIN BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE COBT FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ist.sib.ch/unicoln/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AF010496; AAC16173.1; -.
 DR EMBL: Z46611; CAA86577.1; -.
 DR NCBI: Z05553; IPIV.
 DR InterPro: IPR003200; DBL_PRT.
 DR Pfam: PF02277; DBL_PRT; 1.
 KW Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
 KW Cytosyltransferase;
 FT ACT_SITE 303 303 BASE (BY SIMILARITY).
 FT ACT_SITE 49 49 A -> S (IN REF. 2).
 FT CONTACT 66 66 A -> P (IN REF. 2).
 SQ SEQUENCE 334 AA; 34096 MW; 356A74B0A624637 CRC64;

CC -1- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
 CC -1- CAUTION: THE PLASMID pTIA66 CARRIES TWO T-REGIONS, THE TA AND
 CC TB REGION, BOTH OF WHICH HAVE AN IAAS GENE, WITH LOW HOMOLOGY
 CC BETWEEN THEM. ONLY THE TB-IAAS GENE SEEMS TO BE FUNCTIONAL.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements/>
 CC or send an email to license@isb-sib.ch).
 CC

CC EMBL: X56185; CAA39649.1; ALT_SEQ
 CC PIP: S15452; HJAG1
 CC PIF: S15961; S15901
 CC InterPro: IPR000120, Amidase
 CC Pfam: PF01425; Amidase; 1.
 CC PROSITE: PS00571; AMIDASES; 1.
 CC Hydrolase; Auxin biosynthesis; Crown gall tumor; T-DNA; Plasmid.
 CC SEQUENCE: 467 AA, 49441 MW, 84804MAPS9A1700 CP564;
 CC

Query Match: 75.0%, Score 33, DB 1, Length 467,
 Best Local Similarity 87.5%; Pct Id 66;
 Matches 7; Conservative 6; Mismatches 1; Indels 0, Gaps 0;

QY 1 RLAAALCG 8
 II IIIII
 Db 175 RLPAALCG 182

RESULT 14
 HYIN_AGRT4
 ID HYIN_AGRT4 STANDARD: PPT: 467 AA
 AC P03668
 DT 23-JUL-1986 (Ref 01, Created)
 DT 21-JUL-1986 (Ref 01, Last sequence update)
 DT 01-MAP-2002 (Ref 41, Last annotation update)
 DE Indoleacetamide hydrolase (EC 3.5.1.-) (IAH) (indole-3-acetamide
 DE hydrolase).
 GN TMS2 GP IAAH
 OS Agrobacterium tumefaciens (strain Ach5), and
 OS Agrobacterium tumefaciens.
 CC Plasmid pTIA66, plasmid pTIA66, and plasmid pTIA66.
 CC P-TIA66, P-TIA66, alpha subdivision, Rhizobiaceae group;
 CC Rhizobiaceae; Rhizobium.
 CC NCBI_TaxID=176298, 358;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC PLASMID-pTIA66;
 CC MEDLINE-84144041; PubMed-6366736;
 CC Sciaky D., Thomashow M.F.;
 CC "The sequence of the tms transcript 2 locus of the A. tumefaciens
 CC plasmid pTIA6 and characterization of the mutation in pTIA66 that is
 CC responsible for auxin attenuation";
 CC Nucleic Acids Res. 12:1447-1461(1984).
 CC [2]
 CC SEQUENCE FROM N.A.
 CC PLASMID-pTII5955;
 CC Rarker R.F., Idler K.R., Thompson D.V., Kemp J.D.;
 CC "Nucleotide sequence of the T-DNA region from the Agrobacterium
 CC tumefaciens octopine Ti plasmid pTII5955";
 CC Plant Mol. Biol. 2:335-350(1984)
 CC [3]
 CC SEQUENCE FROM N.A
 CC PLASMID-pTIAch5;
 CC MEDLINE-84207942; PubMed-6327292;
 CC Gielen T., de Boeckeler M., Seufferlein J., Deboeck F., de Greve H.,
 CC Lemmers M., van Montagu M., Schell J.;
 CC "The complete nucleotide sequence of the T-DNA of the Agrobacterium
 CC tumefaciens plasmid pTIAch5";
 CC EMBO J 3 845-846(1984)

FN SEQUENCE FROM N.A.
 RP PLASMID-pTIA66;
 RC MEDLINE-84170374; PubMed-6584906;
 RX Rice H., Montoya A., Hordyski F., Liehlenshein C., Garfinkel D.,
 EA Fuller S., Flores E., Peschen G., Rester E., Gordon M.;
 KI "Nucleotide sequence of the tms genes of the pTIA66 octopine Ti
 EI plasmid: two gene products involved in plant tumorigenesis";
 FI Proc Natl Acad Sci U S A 81:1728-1732(1984).
 FN SEQUENCE FROM N.A.
 RP PLASMID-pTIAch5;
 RC Winans S.C., Zhu J., Oger P.M., Schrammelfen B., Hee/kaas P.J.,
 EA Partand S.K.;
 KI "Octopine-type Ti plasmid sequence";
 FI Submitted (MAP-2000) to the EMBL/Genbank/DDBJ databases.
 CC -1- CATALYTIC ACTIVITY: INDOLE-3-ACETAMIDE (IAM) -> INDOLE-3-
 CC ACETIC ACID (IAA).
 CC -1- PATHWAY: SECOND STEP IN THE BIOSYNTHESIS OF AUXINS FROM
 CC TRYPTOPHAN.
 CC -1- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
 CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements/>
 CC or send an email to license@isb-sib.ch).
 CC

CC EMBL: X00493; CAA25166.1;
 CC PIR: X00493; CAA25166.1;
 CC EMBL: AP243881; AAF77122.1;
 CC EMBL: K02553; AAA92549.1;
 CC PIP: A04501; Q2AGAT.
 CC PIP: K00666; K00666.
 CC PIP: S05886; S05886.
 CC PIF: P01425; Amidase; 1.
 CC Pfam: PF01425; Amidase; 1.
 CC PROSITE: PS00571; AMIDASES; 1.
 CC Hydrolase; Auxin biosynthesis; Crown gall tumor; T-DNA; Plasmid.
 CC SEQUENCE: 467 AA, 49441 MW, 84804MAPS9A1700 CP564;
 CC

Query Match: 75.0%, Score 33, DB 1, Length 467,
 Best Local Similarity 87.5%; Pct Id 66;
 Matches 7; Conservative 6; Mismatches 1; Indels 0, Gaps 0;

QY 1 RLAAALCG 8
 II IIIII
 Db 175 RLPAALCG 182

RESULT 15
 RSEC_ECOLI
 ID RSEC_ECOLI STANDARD: PPT: 159 AA.
 AC P46187;
 DT 01-NOV-1995 (Ref 32, Created)
 DT 01-NOV-1995 (Ref 32, Last sequence update)
 DT 16-OCT-2001 (Ref 40, Last annotation update)
 DE Sigma-E factor regulatory protein rsec.
 GN RSEC OR R2570.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria, gamma subdivision, Enterobacteriaceae;
 CC Escherichia.
 CC NCBI_TaxID=562;
 CC [1]
 CC SEQUENCE FROM N.A
 CC STRAIN-K12 / W3110;
 CC Missiakos D., Raina S.;
 CC Submitted (SEP-1995) to the EMBL/Genbank/DDBJ databases.
 CC [2]
 CC SEQUENCE FROM N.A.
 CC

```

RC STRAIN-K12;
RA Connolly L., de Las Penas A., Gross C.A.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RA Nashimoto H., Saito N.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE:97426617; PubMed:9278503;
RA Blattner F.R.; Plunkett G. III; Bloch C.A.; Perna N.T.; Burland V.;
RA Riley M.; Collado-Vides J.; Glasner J.B.; Rode C.K.; Mayhew G.F.;
RA Gregor J.; Davis N.W.; Kirkpatrick H.A.; Goeden M.A.; Rose D.J.;
RA Mau B.; Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -!- FUNCTION: SEEMS TO MODULATE THE ACTIVITY OF RPOE (SIGMA-E).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (potential).
CC -!- SIMILARITY: TO H.INFLUENZAE H1589 AND H1850.
CC -----
CC This SWISS-prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U37089; AAC45317.1; -;
DR EMBL: U37455; AAC45320.1; -;
DR EMBL: D64044; BAA10917.1; -;
DR EMBL: AF000343; AAC75623.1; -;
DR EcoGene; EG13178; rsec.
KW Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 80 100 POTENTIAL.
FT TRANSMEM 101 121 POTENTIAL.
SQ SEQUENCE 159 AA; 1b659 MW; BCFD8970DF6z77C0 CRC64;

```

```

Query Match      72.7%; Score 32; DB 1; Length 159;
Best Local Similarity 85.7%; Pred.No. 39;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 3 LAALGGA 9
Db 104 VAALGGA 110

```

```

Search completed. September 5, 2002, 15:31:44
Job time: 479 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 GenCore Ltd

OM protein - protein search, using sw model

Run on: September 5, 2002, 15:21:07 : Search time 122.86 seconds
(without alignments)
12,573 Million cell updates/sec

Title: US-09-744-804-41
Perfect score: 44
Sequence: 1 FLNLAALGNA G

Scoring table: BLASTM62
Gapop 10.0, gapext 0.5

Searched: 562222 seqs, 17294929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: SPTREMBL19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_plague.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total sequence distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	100.0	335	4 Q9BTL9	Q9BTL9 homo sapien
2	37	84.1	426	11 Q9WTS3	Q9WTS3 mus musculu
3	37	84.1	463	11 Q9RIX9	Q9RIX9 mus musculu
4	36	81.8	143	16 Q9HX34	Q9HX34 pseudomoxas
5	36	81.8	355	7 Q56601	Q56601 gallus gall
6	36	81.8	355	7 Q31400	Q31400 gallus gall
7	36	81.8	355	7 Q46788	Q46788 gallus gall
8	36	81.8	355	7 Q46789	Q46789 gallus gall
9	36	81.8	355	7 Q46790	Q46790 gallus gall
10	36	81.8	355	7 Q46791	Q46791 gallus gall
11	36	81.8	355	13 Q73901	Q73901 gallus gall
12	36	81.8	355	13 Q73904	Q73904 gallus gall
13	36	81.8	1513	4 Q9NXY3	Q9NXY3 homo sapien
14	35	79.5	230	16 Q9A7Y0	Q9A7Y0 caulebacter
15	35	79.5	611	2 Q910K1	Q910K1 streptomyce
16	35	79.5	818	2 Q93H51	Q93H51 streptomyce

17	34	77.3	171	16 Q9ABJ1	Q9ABJ1 caulebacter
18	34	77.3	312	2 Q93B23	Q93B23 streptomyce
19	34	77.3	375	16 Q9WVS5	Q9WVS5 deinococcus
20	34	77.3	455	5 Q9VQD2	Q9VQD2 drsophila
21	34	77.3	573	2 Q94RY6	Q94RY6 myxococcus
22	34	77.3	879	4 Q9NXX9	Q9NXX9 homo sapien
23	34	77.3	879	4 Q96BC3	Q96BC3 homo sapien
24	34	77.3	1699	4 Q14160	Q14160 homo sapien
25	34	77.3	1744	4 Q901P5	Q901P5 homo sapien
26	34	77.3	1744	1 Q91F35	Q91F35 homo sapien
27	34	77.3	1744	4 Q14305	Q14305 homo sapien
28	33	75.0	28	11 Q9WTS1	Q9WTS1 rattus norv
29	33	75.0	29	11 Q9JLR6	Q9JLR6 rattus norv
30	33	75.0	218	16 Q97225	Q97225 mycobacteri
31	33	75.0	235	10 Q9SM45	Q9SM45 atubidopsis
32	33	75.0	294	2 Q9PSD6	Q9PSD6 arabideteri
33	33	75.0	307	5 Q18327	Q18327 caenorhabdi
34	33	75.0	313	10 Q9S871	Q9S871 atubidopsis
35	33	75.0	334	2 Q93EV5	Q93EV5 streptomyce
36	33	75.0	347	2 Q9RL31	Q9RL31 streptomyce
37	33	75.0	451	2 Q93892	Q93892 carlinocorax
38	33	75.0	466	2 Q9WNR5	Q9WNR5 arabideteri
39	33	75.0	467	2 Q9R695	Q9R695 arabideteri
40	33	75.0	467	2 Q9P712	Q9P712 arabideteri
41	33	75.0	467	2 Q9WNE2	Q9WNE2 arabideteri
42	33	75.0	467	2 Q94205	Q94205 arabideteri
43	33	75.0	469	2 Q9X903	Q9X903 streptomyce
44	33	75.0	472	3 Q59907	Q59907 neurospora
45	33	75.0	473	2 Q93117	Q93117 burkholderi

ALIGNMENTS

RESULT 1

ID Q9BTL9 PRELIMINARY: PRT: 345 AA.

AC Q9BTL9; DT 01-JUN-2001 (11EMBHLrel, 17, Created)

DT 01-JUN-2001 (11EMBHLrel, 17, last sequence update)

DT 01-DEC-2001 (11EMBHLrel, 19, last annotation update)

DE SIMILAR TO MILK FAT GLOBULE-EGF FACTOR 8 PROTEIN.

OS Homo sapiens (Human).

OC Eukaryota, Metazoa, Chordata, Gracilata, Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCHI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=MELANOMA.;

RA Strausberg R.;

FI Submitted (1997-2001) to the Simple Sequence Repeat databases.

CC -1- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.

DR PMR: RC003610; AAH03610.1; -.

DR HSP: P08709; IRF9.

DE InterPro: IPR009561; EGF like.

DR InterPro: IPR001438; EGF_11.

DR InterPro: IPR000421; FA58_C.

DR PRINTS: PR00010; EGFBL000.

DR SMART: SM00181; EGF_1.

DR SMART: SM00001; EGF_Like; 1.

DE SMART: SM00241; FA58C; 2.

OP PROSITE: PS00027; EGF_1; UNKNWN_1

DR PROSITE: PS01186; EGF_2; 1.

DR PROSITE: PS01285; FA58C.1; 1.

DR PROSITE: PS01286; FA58C.2; 2.

KW EGF like domain; Glycoprotein.

SQ SEQUENCE 345 AA: 37523 MW: 23E84E5DE74E9E25 CRK4;

Query Match 100.0%; Score 44; DB 4; Length 335;
Best Local Similarity 100.0%; Pred No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 RLAAALCG 9
DB 5 RLAAALCG 13

RESULT 2
QWTS4 PRELIMINARY: PPT: 426 AA.
AC QWTS4:
DT 01-NOV-1999 (TREMblrel. 13, created)
DT 01-NOV-1999 (TREMblrel. 12, last sequence update)
DT 01-DEC-2001 (TREMblrel. 14, last annotation update)
DE MILK FAT GLOBULE GLYCOPROTEIN MFG-E8 S (SIMILAR TO MILK FAT GLOBULE-EGF FACTOR 8 PROTEIN).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=MAMMARY GLAND;
RX MEDLINE=99120894; PubMed 9920772;
RA Oshima K., Aoki N., Negi M., Kishi M., Kitajima K., Matsuda T.;
PT "Lactation-dependent expression of an mRNA splice variant with an exon
PT for a multiply O-glycosylated domain of mouse milk fat gl-
RT glycoprotein MFG-E8";
RC Biochem. Biophys. Res. Commun. 254:522-528(1999).
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 426 AA; 47167 MW; R182F4A20629882A CPO64;

Query Match 84 18; Score 37; DB 11; Length 426;
Best Local Similarity 87 58; Pred. No. 40;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLAAALCG 8
DB 5 RLAAALCG 12

RESULT 3
QWTS4 PRELIMINARY: PPT: 453 AA.
AC QWTS4:
DT 01-MAY-2000 (TREMblrel. 13, created)
DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
DT 01-DEC-2001 (TREMblrel. 14, last annotation update)
DE MILK FAT GLOBULE GLYCOPROTEIN MFG-E8 LONG FORM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=MAMMARY GLAND;
RX MEDLINE=99120894; PubMed 9920772;
RA Oshima K., Aoki N., Negi M., Kishi M., Kitajima K., Matsuda T.;
PT "Lactation-dependent expression of an mRNA splice variant with an exon
PT for a multiply O-glycosylated domain of mouse milk fat gl-
RT glycoprotein MFG-E8";
RC Biochem. Biophys. Res. Commun. 254:522-528(1999).
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 453 AA; 47167 MW; R182F4A20629882A CPO64;

Query Match 84 18; Score 37; DB 11; Length 426;
Best Local Similarity 87 58; Pred. No. 40;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=MAMMARY GLAND;
RX MEDLINE=99120894; PubMed 9920772;
RA Oshima K., Aoki N., Negi M., Kishi M., Kitajima K., Matsuda T.;
PT "Lactation-dependent expression of an mRNA splice variant with an exon
PT for a multiply O-glycosylated domain of mouse milk fat globule
RT glycoprotein MFG-E8";
RC Biochem. Biophys. Res. Commun. 254:522-528(1999).
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 453 AA; 51269 MW; 0719F09D909F6427 CPO64;

Query Match 84 18; Score 37; DB 11; Length 463;
Best Local Similarity 87 58; Pred. No. 44;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLAAALCG 8
DB 5 RLAAALCG 12

RESULT 4
QWTS4 PRELIMINARY: PPT: 143 AA.
AC QWTS4:
DT 01-MAY-2001 (TREMblrel. 16, created)
DT 01-MAY-2001 (TREMblrel. 16, last sequence update)
DT 01-DEC-2001 (TREMblrel. 18, last annotation update)
DE HYPOTHETICAL PROTEIN PA3986.
GN PA3986.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437333; PubMed 10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Riskey M.J., Brinkman F.S.L., Huftangle W.O., Kowalik D.J., Lauron M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Iori S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
PL Nature 405:959-964(2000).
DR FMBL: AE004816; AAG07373.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 143 AA; 15937 MW; 8845B066D3814B03 CPO64;

Query Match 81.84; Score 36; DB 16; Length 143;
Best Local Similarity 77.89; Pred. No. 23;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 LLAALGCA 9
DB 108 QHICALGCA 116
      IIIIIII
RESULT 5
QY5601 095601 EPIIMMART FFT 355 AA.
AC Q95601 Q95601 (TREMELrel. 02, Created)
DT 01-FEB-1997 (TREMELrel. 02, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE MAJOR HISTOCOMPATIBILITY COMPLEX CLASS I GLYCOPROTEIN HAPLOTYPE B*1.
GN H*1V21 OR B-FIV.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
CX [1]
RN SEQUENCE FROM N.A.
PX MEDLINP-95347411; PubMed-7621880;
PA Fulton J.F., Thakor E.L., Rivas J.P., Hunt H.D.:
RT "Functional analysis of avian class I (B*1V) glycoproteins by epitope
RT tagging and mutagenesis in vitro."
KL Eur. J. Immunol. 25:2965-2976 (1995)
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=1515 R1 CONGENIC;
RA Hunt H.D.:
KL Submitted (JAN-1998) to the EMBL/GenBank/EBI Databases.
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
CC IMMUNE SYSTEM (BY SIMILARITY).
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA 2-
CC MICROGLOBULIN) (BY SIMILARITY).
DR EMBL: S78682; AAB34945.1; -.
DR EMBL: AF013493; AAC17590.1; -.
DR HSSP: P13599; 3F8U.
DR InterPro: IPR003597; Iq_c1.
DR InterPro: IPR003006; Iq_MHC.
DR InterPro: IPR001039; MHC_1.
DR Pfam: PF00047; Iq_1.
DR Pfam: PF00129; MHC_1.
DR ProDom: P000050; MHC_1.
DR SMART: SM00407; IqC1; 1.
DR PROSITE: PS00290; Iq_MHC; UNKNOWN_1.
KW Glycoprotein; MHC; Transmembrane_1.
SQ SEQUENCE 355 AA; 38960 MW; A751A09C46D94FE9 PC64;

Query Match 81.8%; Score 36; DB 7; Length 355;
Best Local Similarity 87.5%; Pred. NO. 52;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0.

QY 2 LLAALGCA 9
DB 12 LLAALGCA 19
      IIIIIII
RESULT 6
QY1400 Q31400 PRELIMINARY; PRT; 355 AA.
AC Q31400;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE MAJOR HISTOCOMPATIBILITY COMPLEX CLASS I GLYCOPROTEIN PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;

```

```

RN SEQUENCE FROM N.A.
RC STRAIN=CB; TISSUE=Bursa;
RA Walling H., Avila D., Riepert P., Salomonson J., Vilheis F., Miles M.,
RA Kaufman J.:
RT "Peptide motifs for the dominantly expressed class I molecule of the
RT chicken major histocompatibility complex and resistance to infectious
RT pathogens."
KL Submitted (JAN-1997) to the EMBL/GenBank/EBI Databases.
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
CC IMMUNE SYSTEM (BY SIMILARITY).
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA 2-
CC MICROGLOBULIN) (BY SIMILARITY).
DR EMBL: 254336; CAA91118.1; -.
DR HSSP: P13599; 3F8U.
DR InterPro: IPR003597; Iq_c1.
DR InterPro: IPR003006; Iq_MHC.
DR InterPro: IPR001039; MHC_1.
DR Pfam: PF00047; Iq_1.
DR Pfam: PF00129; MHC_1.
DR ProDom: P000050; MHC_1.
DR SMART: SM00407; IqC1; 1.
DR PROSITE: PS00290; Iq_MHC; UNKNOWN_1.
KW Glycoprotein; signal; Transmembrane.
RT SIGNAL 1 21 POTENTIAL.
RT CHAIN 22 355 MAJOR CLASS I GLYCOPROTEIN.
SQ SEQUENCE 355 AA; 39306 MW; B91987452F1917A PC64;

Query Match 81.8%; Score 36; DB 7; Length 355;
Best Local Similarity 87.5%; Pred. NO. 52;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLAALGCA 9
DB 12 LLAALGCA 19
      IIIIIII
RESULT 7
QY6788 Q46788 PRELIMINARY; PRT; 355 AA.
AC Q46788;
DT 01-JUN-1998 (TREMELrel. 06, Created)
DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE MHC CLASS I GLYCOPROTEIN.
GN B-FIV.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_TaxID=9031;
CX [1]
RN SEQUENCE FROM N.A.
RC STRAIN 1515 R5 CONGENIC;
RX MEDLINE 98221119; PubMed 9553152;
RA Hunt H.D., Fulton J.F.:
RT "Analysis of polymorphisms in the major expressed class I locus (B-
RT FIV) of the chicken."
KL Immunogenetics 47:456-467 (1998).
CC -1- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
CC IMMUNE SYSTEM (BY SIMILARITY).
CC -1- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA 2-
CC MICROGLOBULIN) (BY SIMILARITY).
DR EMBL: AF013491; AAC17588.1; -.
DR HSSP: P13599; 3F8U.
DR InterPro: IPR003597; Iq_c1.
DR InterPro: IPR003006; Iq_MHC.
DR InterPro: IPR001039; MHC_1.
DR Pfam: PF00047; Iq_1.
DR Pfam: PF00129; MHC_1.
DR ProDom: P000050; MHC_1.
DR SMART: SM00407; IqC1; 1.

```

DR PROSITE: PS00290; IG_MHC; UNKNOWN_1.
 KW Glycoprotein; Transmembrane.
 SQ SEQUENCE 355 AA; 38453 MW; F7BCF9B88F61582E CRC64;

Query Match 81.8%; Score 36; DB 7; Length 355;
 Best Local Similarity 87.5%; Pred. No. 52;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLAALGGA 9
 II|I|I|I|I|
 DB 12 LLAAVCGA 19

RESULT 8
 046789 PRELIMINARY; PRT; 355 AA
 AC 046789;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE MHC CLASS I GLYCOPROTEIN.
 GN B-FIV.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:1515 B13 CONGENIC;
 RX MEDLINE:98221119; PubMed:9554152;
 RA Hunt H.D., Fulton J.E.;
 RT "Analysis of polymorphisms in the major expressed class I locus (B-FIV) of the chicken."
 RL Immunogenetics 47:456-467(1998).
 CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM (BY SIMILARITY).
 CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN) (BY SIMILARITY).
 DR EMBL: AF013494; AAC17589.1; -.
 DR InterPro: IPR003597; Iq_cl.
 DR InterPro: IPR003006; Iq_MHC.
 DR InterPro: IPR001039; MHC_1.
 DR Pfam: PF00047; Iq; 1.
 DR Pfam: PF00129; MHC_I; 1.
 DR ProDom: PD000050; MHC_I; 1.
 DR SMART: SM00407; IGC1; 1.
 DR PROSITE: PS00290; IG_MHC; UNKNOWN_1.
 KW Glycoprotein; Transmembrane.
 SQ SEQUENCE 355 AA; 39049 MW; 30F7F50714F0F0 CRC64;

Query Match 81.8%; Score 36; DB 7; Length 355;
 Best Local Similarity 87.5%; Pred. No. 52;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLAALGGA 9
 II|I|I|I|I|
 DB 12 LLAAVCGA 19

RESULT 9
 046790 PRELIMINARY; PRT; 355 AA.
 AC 046790;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE MHC CLASS I GLYCOPROTEIN.
 GN B-FIV.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:1515 B13 CONGENIC;
 RX MEDLINE:98221119; PubMed:9554152;
 RA Hunt H.D., Fulton J.E.;
 RT "Analysis of polymorphisms in the major expressed class I locus (B-FIV) of the chicken."
 RL Immunogenetics 47:456-467(1998).
 CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM (BY SIMILARITY).
 CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN) (BY SIMILARITY).
 DR EMBL: AF013494; AAC17591.1; -.
 DR InterPro: IPR003597; Iq_cl.
 DR InterPro: IPR003006; Iq_MHC.
 DR InterPro: IPR001039; MHC_1.
 DR Pfam: PF00047; Iq; 1.
 DR Pfam: PF00129; MHC_I; 1.
 DR ProDom: PD000050; MHC_I; 1.
 DR SMART: SM00407; IGC1; 1.
 DR PROSITE: PS00290; IG_MHC; UNKNOWN_1.
 KW Glycoprotein; Transmembrane.
 SQ SEQUENCE 355 AA; 39398 MW; 8B08F669AFF5AF37 CRC64;

Query Match 81.8%; Score 36; DB 7; Length 355;
 Best Local Similarity 87.5%; Pred. No. 52;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLAALGGA 9
 II|I|I|I|I|
 DB 12 LLAAVCGA 19

RESULT 10
 046791 PRELIMINARY; PRT; 355 AA.
 AC 046791;
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE MHC CLASS I GLYCOPROTEIN.
 GN B-FIV.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:B17 CONGENIC;
 RX MEDLINE:98221119; PubMed:9554152;
 RA Hunt H.D., Fulton J.E.;
 RT "Analysis of polymorphisms in the major expressed class I locus (B-FIV) of the chicken."
 RL Immunogenetics 47:456-467(1998).
 CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE IMMUNE SYSTEM (BY SIMILARITY).
 CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-MICROGLOBULIN) (BY SIMILARITY).
 DR EMBL: AF013495; AAC17592.1; -.
 DR HSSP: P13599; 3FRU.
 DR InterPro: IPR003597; Iq_cl.
 DR InterPro: IPR003006; Iq_MHC.
 DR InterPro: IPR001039; MHC_1.
 DR Pfam: PF00047; Iq; 1.
 DR Pfam: PF00129; MHC_I; 1.
 DR ProDom: PD000050; MHC_I; 1.
 DR SMART: SM00407; IGC1; 1.
 DR PROSITE: PS00290; IG_MHC; UNKNOWN_1.


```

KW Glycoprotein; Transmembrane.
SQ SEQUENCE 355 AA; 34170 MW; 563825775645AAAH (P=1.4);

Query Match      81.8%; Score 36; DB 7; Length 355;
Best Local Similarity 87.5%; Pred. No. 52;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLAALCGA 9
DB 12 LLAAVCGA 19

RESULT 11
073901
AC 073901 PRELIMINARY; PRT; 355 AA.
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE B LOCUS F ALPHA CHAIN 2.
GN BFA2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=89030642; PubMed=341149;
RA Guillemot F., Billaut A., Pourquie O., Behar G., Chausse A.M.,
RA Zoorob R., Kreibich G., Auffray G.;
RT "A molecular map of the chicken major histocompatibility complex; the
RT class II beta genes are closely linked to the class I genes and the
RT nuclear organizer."
RL EMBO J. 7:2775-2785(1988).
RN 12
RP SEQUENCE FROM N.A.
RA Milne S., Kaufman J., Beck S.;
RT "DNA sequencing and analysis of the chicken major histocompatibility
RT complex."
RX PubMed=341149; PubMed=341149;
CC -!- FUNCTION: INVOLVED IN THE PRESENTATION OF FOREIGN ANTIGENS TO THE
CC IMMUNE SYSTEM (BY SIMILARITY).
CC -!- SUBUNIT: DIMER OF ALPHA CHAIN AND A BETA CHAIN (BETA-2-
CC MICROGLOBULIN) (BY SIMILARITY).
DR EMBL: AL023516; CAA18969.1; -.
DR HSSP: P13599; 3FEU.
DR InterPro: IPR003597; I4_c1.
DR GeneID: 24603006; I4_MHC.
DR InterPro: IPR001039; MHC_1.
DR Pfam: PF00047; Iq; 1.
DR ProDom: PD000050; MHC_1; 1.
DR SMART: SM00407; IGH1; 1.
DR PROSITE: PS00290; IGH_MHC; UNKNOWN_1.
KW Glycoprotein; Transmembrane.
SQ SEQUENCE 355 AA; 34200 MW; 080987452F18467A CRC64;

Query Match      81.8%; Score 36; DB 13; Length 355;
Best Local Similarity 87.5%; Pred. No. 52;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLAALCGA 9
DB 12 LLAAVCGA 19

RESULT 13
09NXY3
AC 09NXY3 PRELIMINARY; PRT; 353 AA.
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE FLJ00002 PROTEIN (FRAGMENT).
GN FLJ00002.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RX TISSUE=SPLEEN;
RA Chara O., Kifuno R., Kagase T., Kamura K.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
RT spleen."
RL Submitted (118 1999) to the EMBL, GenBank, DDBJ databases.
DR EMBL: AK000002; BA02227.1; -.

```

DR HSP: P14569; INBD.
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR001140; ABC_transporter_tunh.
 DR InterPro: IPR003439; ABC_transporter.
 DR InterPro: IPR001687; ATP_GTP_A.
 DR Pfam: PF00664; ABC_membrane; 2.
 DR Pfam: PF00005; ABC_tran; 2.
 DR SMART: SM00382; AAA; 2.
 DR PROSITE: PS00211; ABC_TRANSPORTER; DRYN-WNL;
 KW ATP-binding.
 FT NON_TER 1
 SQ SEQUENCE 1513 AA; 163831 MW; FBA7AARDE5AE3C CRC64;

Query Match 81.8%; Score 36; DR 4; Length 1513;
 Best local Similarity 77.6%; Pred. No. 2e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLAALCGA 9
 |||||
 DB 24 LLAALCGS 32
 |||||
 RESULT 14
 QYAYO PRELIMINARY; PRI; 230 AA.
 AC QYAYO:
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE HYDROTHERMAL PROTEIN CCI589.
 GN CCI589.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 OR Caulobacter.
 OX NCBI_TaxID 69394;
 RN [1]
 KC SEQUENCE FROM N.A.
 RC STRAIN ATCC 19089 / CH15;
 KK MEDLINE-31173598; PubMed 11259647;
 KA Eisen J.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 KA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Fly R.,
 KA DeRoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 KA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 KA Uitterlinden T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 KA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.; "
 RT "Complete genome sequence of Caulobacter crescentus.";
 RL Proc. Natl. Acad. Sci. U S A 98:4116-4141(2001)
 DR EMBL: AF005833; AAK23568.1; "
 DR TIGR: CCI589; "
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 230 AA; 24454 MW; DEBACD3F28E54B01 CRC64;

Query Match 79.5%; Score 35; DR 16; Length 230;
 Best local Similarity 87.5%; Pred. No. 54;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLAALCGA 9
 |||||
 DB 6 LLAALCGA 13
 |||||
 RESULT 15
 QYLOKI PRELIMINARY; PRI; 611 AA.
 AC QYLOKI:
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PUTATIVE MEMBRANE PROTEIN.
 GN SCDA0A.10C.

OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomyces, Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 KC SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 KA Sanger K.J., Harris D.;
 FT Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 KC SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 KA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
 FT Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 KC SEQUENCE FROM N.A.
 RC STRAIN=A3(2);
 KK MEDLINE-97000351; PubMed 8843436;
 KA Rodenbach M., Kieser R.M., Denapate D., Eichner A., Cullum J.,
 KA Kinashi H., Hopwood D.A.;
 RT "A set of ordered cosmids and a detailed genetic and physical map for
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL161691; CAB81855.1; "
 DR InterPro: IPR003838; DUF214.
 DR Pfam: PF02697; DUF214; 1.
 SQ SEQUENCE 611 AA; 61744 MW; 8F959720773AEDF9 CRC64;

Query Match 79.5%; Score 35; DR 2; Length 611;
 Best local Similarity 87.5%; Pred. No. 1.3e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLAALCGA 9
 |||||
 DB 140 LLAALCGA 147
 |||||

Search completed: September 5, 2002, 15:31:08
 Job time: 473 sec

